Internet Appendix

A Proofs

Lemma 1. *Letting* Σ *be defined as in A3, (i)* $\Sigma \in \mathbb{R}$*, (ii)* $\frac{\Sigma}{2} \leq \theta$*, and (iii)* $\frac{\Sigma}{2} \geq \frac{X\theta\lambda_j}{\lambda_i}$*, and (iv) if* $X \geq 2$*, then* $\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i} > 0$.

Proof of lemma 1. To prove the first claim, notice that by A1,

$$\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i} \geq \theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta}{X^2} = \left(\theta - \frac{4\alpha}{X^2}\right)^2 \geq 0,$$

which therefore implies that $\Sigma \in \mathbb{R}$. The proof of the second claim has two parts. First, if $X = 1$, then the result follows directly from A1. Second, if $X \geq 2$, then by A1,

$$\theta + \frac{4\alpha}{X^2} - \sqrt{\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i}} \leq \theta + \frac{4\alpha}{X^2} - \sqrt{\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta}{X^2}} = \theta + \frac{4\alpha}{X^2} - \left|\theta - \frac{4\alpha}{X^2}\right| \leq 2\theta.$$

To prove the third claim, note that by plugging A2 into A3, we must have $\lambda_i\left(1 - \frac{1}{\theta}\frac{\Sigma}{2}\right)\frac{\Sigma}{2} > \lambda_j X\left(\theta - \frac{\Sigma}{2}\right)$, which can be manipulated to yield $\left(\frac{\Sigma}{2} - \theta\right)\left(\frac{\Sigma}{2} - \frac{X\theta\lambda_j}{\lambda_i}\right) < 0$. In light of the second claim, the first factor must be negative, which implies that the second factor must be positive. In light of the proof of the first claim, proving the third claim reduces to deriving a contradiction from the combination of $X \geq 2$ and $\theta = \frac{4\alpha}{X^2}$. If this is the case, then $\Sigma = 2\theta$. Plugging that into A3 yields $0 \geq \lambda_j X(\sigma - \theta)$, which contradicts A2. $\square$

Proof of proposition 1. The proof proceeds in two parts. First we describe equilibrium strategies, and second we show that no player has a profitable deviation.

Part 1 (description): Let s^* be defined as in the proposition: $s^* = \theta\left(1 + \frac{\lambda_j}{\lambda_i}\right)$. The strategy of the exchange is to set a make fee τ^*_{make} and a take fee $\tau^*_{\text{take}} < \frac{s^*}{2}$ such that

$$\tau^*_{\text{make}} + \tau^*_{\text{take}} = \frac{\lambda_i\left(1 - \frac{1}{\theta}\frac{s^*}{2}\right)\frac{s^*}{2} - \lambda_j\left(\sigma - \frac{s^*}{2}\right)}{\lambda_i\left(1 - \frac{1}{\theta}\frac{s^*}{2}\right) + \lambda_j}.$$

Note that A3 ensures this total fee will be nonnegative. We do not specify the individual fees, as only the total fee will be relevant for the subsequent analysis.[36]

[36] In other words, equilibrium pins down only the total fee $\tau^*_{\text{make}} + \tau^*_{\text{take}}$, but not the exact decomposition

An investor who arrives at time t with type $(\tilde{l},\tilde{\theta})$ behaves as specified: choosing a quantity $y \in \{-1,0,1\}$ to maximize $\hat{u}_t(y|\tilde{l},\tilde{\theta})$. Thus, it remains to specify the strategies of the HFTs. One HFT plays the role of "liquidity provider." A second HFT plays the role of "enforcer." The remaining HFTs (infinitely many) play the role of "snipers."

To define the strategy of the liquidity provider, we consider the following polynomial in s, where we use $\tau = (\tau_{\text{make}},\tau_{\text{take}})$ to denote the fees set by the exchange:

$$\pi(s|\tau) \coloneqq \lambda_i \left(1 - \frac{1}{\theta}\frac{s}{2}\right)\left(\frac{s}{2} - \tau_{\text{make}} - \tau_{\text{take}}\right) + \lambda_j \left(\frac{s}{2} - \tau_{\text{make}} - \tau_{\text{take}} - \sigma\right).$$

The polynomial $\pi(s|\tau)$ represents, roughly speaking, the profits that would accrue to the liquidity provider if the spread were s and trading fees were given by τ. There are two cases. First, if τ is such that no value of $s \in [0,2\theta]$ is a root of $\pi(s|\tau)$, then the liquidity provider never quotes. Otherwise, let $s(\tau)$ be defined implicitly as the smallest such root.[37] In that case, the liquidity provider acts as follows. At time zero, she submits to the exchange a limit order to buy one share at $\hat{b}_0 = v_0 - \frac{s(\tau)}{2} + \tau_{\text{take}}$ and a limit order to sell one share at $\hat{a}_0 = v_0 + \frac{s(\tau)}{2} - \tau_{\text{take}}$.[38] If one of her standing limit orders is filled at a time when v_t has not jumped, then she immediately submits an identical order to replace it. If v_t jumps, then she immediately submits to the exchange the following orders: (i) cancellations for her limit orders, (ii) a limit order to buy one share at $\hat{b}_{t^+} = v_{t^+} - \frac{s(\tau)}{2} + \tau_{\text{take}}$, and (iii) a limit order to sell one share at $\hat{a}_{t^+} = v_{t^+} + \frac{\hat{s}(\tau)}{2} - \tau_{\text{take}}$.[39]

The strategy of the enforcer is as follows. She never submits any orders unless the liquidity provider is observed to have deviated, in which case the enforcer begins to take the actions that were prescribed for the liquidity provider.

The strategy of a sniper is as follows. If v_t jumps upward (downward), then she immediately submits to the exchange an immediate-or-cancel order to buy (sell) at the price $v_{t^-} + \sigma - \tau_{\text{take}}$ $(v_{t^-} - \sigma + \tau_{\text{take}})$.

Part 2 (verification): It can be shown that $s(\tau^*) = s^*$. We now argue that if all other players behave as specified, then the liquidity provider has no profitable deviations. The arguments are similar to those in Budish et al. (2015, proof of Proposition 1). By lemma 1(ii), $s^*/2 \le \theta$, and by A2, $s^*/2 < \sigma$. Thus, if the liquidity provider sets a spread $s \le s^*$, then her flow profits are $\pi(s|\tau^*)$. These profits are zero at s^*, since $s(\tau^*)$ is defined as a root of $\pi(s|\tau^*)$, in particular the smallest root. Moreover, since $\pi(s|\tau^*)$ is a concave, second-order polynomial, profits must be negative at spreads $s < s^*$. Thus, it is not profitable to deviate by setting

thereof. Thus, a variety of fee structures are consistent with the model, including: (i) those in which both sides pay a fee, as is the case in Australia, (ii) "maker-taker" fee structures in which makers receive a rebate while takers pay a fee and (iii) "inverted" fee structures in which takers receive a rebate while makers pay a fee. The same observation applies to the oligopoly equilibrium characterized by proposition 2. The decomposition cannot, however, be completely arbitrary: it is infeasible to have a take fee that exceeds half the cum-fee spread (for then the quoted spread would be negative). This is the reason for requiring $\tau^*_{\text{take}} < \frac{s^*}{2}$.

[37] Thus, $s(\tau) = \tau_{\text{make}} + \tau_{\text{take}} + \theta\left(1 + \frac{\lambda_j}{\lambda_i}\right) - \sqrt{(\tau_{\text{make}} + \tau_{\text{take}})^2 - 2\theta(\tau_{\text{make}} + \tau_{\text{take}})\left(1 + \frac{\lambda_j}{\lambda_i}\right) + 2\theta\frac{\lambda_j}{\lambda_i}(\theta - 2\sigma) + \theta^2\left(1 + \frac{\lambda_j^2}{\lambda_i^2}\right)}$.

[38] The resulting cum-fee spread is $s(\tau)$. It depends on the make and take fees only through their sum.

[39] For any continuous time variable X_t, we use X_{t^+} to denote $\lim_{s\to t^+} X_s$ and X_{t^-} to denote $\lim_{s\to t^-} X_s$.

a smaller spread. It is also not profitable to deviate by setting a larger spread, since the enforcer would then undercut her, and she would receive none of the benefits (from investor orders), but might receive adverse selection costs (from sniper orders). Finally, it is also not profitable to deviate by quoting more than a single unit at either the bid or the ask, since her benefits would be the same (only one unit at each is needed to satisfy investor demand) but her costs would increase (since more units are exposed to adverse selection from snipers).

We now argue that the snipers and the enforcer have no profitable deviations. The arguments are also similar to those in Budish et al. (2015, proof of Proposition 1). They also earn zero profits in the equilibrium, and it therefore remains to show that none of them possesses a deviation that would yield positive profits. It is also not profitable to attempt to provide liquidity at a smaller spread than the liquidity provider, since that would result in negative expected profits for the same reason as above. It is also not profitable to attempt to provide liquidity at the same spread as the liquidity provider, since these quotes have the same adverse selection costs (from sniper orders) that the liquidity provider faces in equilibrium but only half the benefits (from investor orders), and would therefore result in negative expected profits. Finally, it is not profitable to attempt to provide liquidity at a larger spread than the liquidity provider, since these orders would receive none of the benefits (from investor orders), but might receive adverse selection costs (from sniper orders).

We finally argue that the exchange has no profitable deviations. Given the behavior of the traders, the profits of the exchange are zero for the case in which the liquidity provider does not quote. In the other case, the profits of the exchange are

$$(\tau_{\text{make}} + \tau_{\text{take}})\left[\lambda_j + \lambda_i\left(1 - \frac{1}{\theta}\frac{s(\tau)}{2}\right)\right],$$

which, using the fact that $s(\tau)$ is a root of $\pi(s|\tau)$, can be shown to equal

$$\lambda_i\left(1 - \frac{1}{\theta}\frac{s(\tau)}{2}\right)\frac{s(\tau)}{2} + \lambda_j\left(\frac{s(\tau)}{2} - \sigma\right).$$

The above expression is a concave function of the spread, $s(\tau)$, which is maximized when $s(\tau) = s^*$. Since $s(\tau^*) = s^*$, the exchange has no profitable deviations to other fee structures under which the liquidity provider quotes. Furthermore, by A3, this yields nonnegative profits for the exchange, so the exchange also has no profitable deviations to fee structures for which the liquidity provider does not quote. □

Proof of proposition 2. The proof proceeds in two parts. First we describe equilibrium strategies, and second we show that no player has a profitable deviation.

Part 1 (description): Let s^* be defined as in the proposition: $s^* = \theta + \frac{4\alpha}{X^2} - \sqrt{\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i}}$. The strategy of each exchange x is to set a make fee $\tau^*_{x,\text{make}}$ and a take fee $\tau^*_{x,\text{take}} < \frac{s^*}{2}$ such that

$$\tau^*_{x,\text{make}} + \tau^*_{x,\text{take}} = \frac{\lambda_i\left(1 - \frac{1}{\theta}\frac{s^*}{2}\right)\frac{s^*}{2} - \lambda_j X\left(\sigma - \frac{s^*}{2}\right)}{\lambda_i\left(1 - \frac{1}{\theta}\frac{s^*}{2}\right) + \lambda_j X}.$$

Note that A3 ensures this total fee will be nonnegative. We do not specify the individual fees, as only the total fee will be relevant for the subsequent analysis.

An investor who arrives at time t with type $(\tilde{l},\tilde{\theta})$ behaves as specified: choosing an exchange $x \in \{1,\dots,X\}$ and a quantity $y \in \{-1,0,1\}$ to maximize $\hat{u}_t(y,x|\tilde{l},\tilde{\theta})$. Thus, it remains to specify the strategies of the HFTs, which will be merely the multi-exchange analogues of the monopoly case of proposition 1. One HFT per exchange plays the role of "liquidity provider." A second HFT plays the role of "enforcer." The remaining HFTs (infinitely many) play the role of "snipers."

To define the strategy of the liquidity provider for exchange x, we consider the following polynomial in s_x, where we use $\tau_x = (\tau_{x,\text{make}},\tau_{x,\text{take}})$ to denote the fees set by exchange x.

$$\pi(s_x|\tau_x) := \lambda_i\left[\frac{X}{2\alpha}\left(\frac{s^*}{2}-\frac{s_x}{2}\right)+\frac{1}{X}\right]\left(1-\frac{1}{\theta}\frac{s_x}{2}\right)\left(\frac{s_x}{2}-\tau_{x,\text{make}}-\tau_{x,\text{take}}\right)+\lambda_j\left(\frac{s_x}{2}-\tau_{x,\text{make}}-\tau_{x,\text{take}}-\sigma\right).$$

The polynomial $\pi(s_x|\tau_x)$ represents, roughly speaking, the profits that would accrue to the liquidity provider on exchange x if the spread on that exchange were s_x, trading fees on that exchange were given by τ_x, and if the spreads on other exchanges were s^*. There are two cases. First, if τ_x is such that no value of $s_x \in [0,2\theta]$ is a root of $\pi(s_x|\tau_x)$, then the liquidity provider never quotes. Otherwise, let $s(\tau_x)$ be defined implicitly as the smallest such root. In that case, the liquidity provider for exchange x takes actions analogous to those delineated in the proof of proposition 1 in order to maintain a cum-fee spread on exchange x of $s(\tau_x)$.

The strategy of the enforcer is as follows. She never submits any orders unless a liquidity provider is observed to have deviated, in which case the enforcer begins to take the actions that were prescribed for the deviating liquidity provider.

The strategy of a sniper is as follows. If v_t jumps upward (downward), then she immediately submits to each exchange an immediate-or-cancel order to buy (sell) at the price $v_{t^-}+\sigma-\tau_{\text{take}}$ $(v_{t^-}-\sigma+\tau_{\text{take}})$.

Part 2 (verification): We claim that $s(\tau_x^*) = s^*$. It can be shown that s^* is a root of $\pi(s_x|\tau_x^*)$, so it remains only to be shown that it is the smallest root. To begin, note that since $\tau^*_{x,\text{make}}+\tau^*_{x,\text{take}} \geq 0$ and $\sigma > \theta$, it follows that $\pi(2\theta|\tau_x^*) < 0$. In addition, by lemma 1(ii), $s^* \leq 2\theta$. Because $\pi(s_x|\tau_x^*)$ is a third-order polynomial in s_x with a positive leading coefficient, we conclude from these facts that it has three roots and that s^* is not the largest of those. In addition, it can be shown that $\pi'(s^*|\tau_x^*) \geq 0$, which means that s^* cannot be the middle root. This establishes the claim.

We now argue that if all other players behave as specified, then the liquidity provider at exchange x has no profitable deviations. By lemma 1(ii), $s^*/2 \leq \theta$, and by A2, $s^*/2 < \sigma$. Thus, if the liquidity provider at exchange x sets a spread $s_x \leq s^*$, then her flow profits are $\pi(s_x|\tau_x^*)$. These profits are zero at s^*, since $s(\tau^*)$ is defined as a root of $\pi(s_x|\tau_x^*)$, in particular the smallest root. Moreover, since $\pi(s_x|\tau_x^*)$ is a third-order polynomial with a positive leading coefficient, profits must be negative at spreads $s_x < s^*$. Thus, it is not profitable to deviate by setting a smaller spread. That other types of deviations are not profitable can be argued as in the proof of proposition 1. And that the other HFTs also have no profitable deviations can similarly be argued as in the proof of proposition 1.

We now argue that the exchange has no profitable deviations. Given the behavior of the traders and other exchanges, the profits of exchange x are zero for the case in which the liquidity provider at exchange x does not quote. For the case in which the liquidity provider does quote, the profits of exchange x are

$$(\tau_{x,\text{make}} + \tau_{x,\text{take}}) \left(\lambda_j + \lambda_i \left[\frac{X}{2\alpha} \left(\frac{s^*}{2} - \frac{s(\tau_x)}{2} \right) + \frac{1}{X} \right] \left(1 - \frac{1}{\theta} \frac{s(\tau_x)}{2} \right) \right),$$

which, using the fact that $s(\tau_x)$ is a root of $\pi(s_x|\tau_x)$, can be shown to equal

$$\lambda_i \left[\frac{X}{2\alpha} \left(\frac{s^*}{2} - \frac{s(\tau_x)}{2} \right) + \frac{1}{X} \right] \left(1 - \frac{1}{\theta} \frac{s(\tau_x)}{2} \right) \frac{s(\tau_x)}{2} + \lambda_j \left(\frac{s(\tau_x)}{2} - \sigma \right).$$

Note that when the liquidity provider does quote, she sets a spread in the domain $[0, 2\theta]$. We claim that the spread $s(\tau_x) = s^*$ maximizes the above expression on this domain. It can be shown that s^* is a critical point of the expression. Note also that by A3 the expression is nonnegative at s^*. Moreover, the expression is negative at 2θ (the first term in the last factor is zero at 2θ, and the second term is negative, since $\sigma > \theta$ by A2). In addition, by lemma 1(ii), $s^* \leq 2\theta$. Because the expression is a third-order polynomial in $s(\tau_x)$ with a positive leading coefficient, we conclude from these facts that s^* must be the unique local maximum. It therefore remains only to show that the expression is not larger at either endpoint. We have already argued that the expression is nonnegative at s^* and negative at 2θ. It is also negative at 0, which establishes the claim.

Since $s(\tau_x^*) = s^*$, the exchange has no profitable deviations to other fee structures under which the liquidity provider quotes. Furthermore, by A3, this yields nonnegative profits for the exchange, so the exchange also has no profitable deviations to fee structures for which the liquidity provider does not quote. □

Proof of corollary 3. We begin by establishing that $\alpha \geq \frac{\theta\lambda_j}{\lambda_i}$. Suppose to the contrary that $\alpha < \frac{\theta\lambda_j}{\lambda_i}$. Note that this can be the case only if $\frac{\lambda_j}{\lambda_i} > 0$. Then because $s^*_{duopoly}$ is weakly increasing in α (*cf.* proposition 4), we obtain

$$s^*_{duopoly} \leq \theta + \frac{\theta\lambda_j}{\lambda_i} - \sqrt{\theta^2 + \frac{\theta^2\lambda_j^2}{\lambda_i^2} - \frac{4\theta^2\lambda_j^2}{\lambda_i^2}} = \theta \left(1 + \frac{\lambda_j}{\lambda_i} - \sqrt{1 - \frac{3\lambda_j^2}{\lambda_i^2}} \right).$$

On the other hand, lemma 1(iii) implies that $s^*_{duopoly} \geq \frac{4\theta\lambda_j}{\lambda_i}$. These two requirements are consistent with each other only if $\frac{4\theta\lambda_j}{\lambda_i} \leq \theta \left(1 + \frac{\lambda_j}{\lambda_i} - \sqrt{1 - \frac{3\lambda_j^2}{\lambda_i^2}} \right)$, or equivalently,

$$1 - 3\frac{\lambda_j}{\lambda_i} \geq \sqrt{1 - 3\left(\frac{\lambda_j}{\lambda_i}\right)^2}. \tag{12}$$

By A1, we have $\frac{\lambda_j}{\lambda_i} \leq \frac{1}{2}$. And as observed above, we also have $\frac{\lambda_j}{\lambda_i} > 0$. For such values, (12) cannot be satisfied, and we obtain the desired contradiction.

By propositions 1 and 2, $s^*_{monopoly} \le s^*_{duopoly}$ if and only if $\theta\left(1+\frac{\lambda_j}{\lambda_i}\right) \le \theta+\alpha-\sqrt{\theta^2+\alpha^2-\frac{4\alpha\theta\lambda_j}{\lambda_i}}$. Rearranging, we obtain $\alpha-\frac{\theta\lambda_j}{\lambda_i} \ge \sqrt{\theta^2+\alpha^2-\frac{4\alpha\theta\lambda_j}{\lambda_i}}$. We have shown above that $\alpha \ge \frac{\theta\lambda_j}{\lambda_i}$. Thus, we can square both sides to conclude that $s^*_{monopoly} \le s^*_{duopoly}$ if and only if $\alpha^2 - \frac{2\alpha\theta\lambda_j}{\lambda_i} + \frac{\theta^2\lambda_j^2}{\lambda_i^2} \ge \theta^2+\alpha^2-\frac{4\alpha\theta\lambda_j}{\lambda_i}$, which, when rearranged, yields the desired condition. □

Proof of proposition 4. In the case of a monopoly ($X = 1$), the claims follow straightforwardly from the derivatives of the expression for s^* given in proposition 1 with respect to those parameters. We therefore focus below on the case of an oligopoly ($X \ge 2$). In this case, the claims follow from the derivatives of the expression for s^* given in proposition 2 with respect to those parameters. To establish this, we first compute these derivatives:

$$\frac{\partial s^*}{\partial \lambda_i} = \frac{-4\alpha\theta\lambda_j}{\lambda_i^2 X\sqrt{\theta^2+\frac{16\alpha^2}{X^4}-\frac{8\alpha\theta\lambda_j}{X\lambda_i}}} \qquad \frac{\partial s^*}{\partial \lambda_j} = \frac{4\alpha\theta}{\lambda_i X\sqrt{\theta^2+\frac{16\alpha^2}{X^4}-\frac{8\alpha\theta\lambda_j}{X\lambda_i}}}$$

$$\frac{\partial s^*}{\partial \alpha} = \frac{\frac{4\theta\lambda_j}{X\lambda_i}-\frac{16\alpha}{X^4}+\frac{4}{X^2}\sqrt{\theta^2+\frac{16\alpha^2}{X^4}-\frac{8\alpha\theta\lambda_j}{X\lambda_i}}}{\sqrt{\theta^2+\frac{16\alpha^2}{X^4}-\frac{8\alpha\theta\lambda_j}{X\lambda_i}}}$$

The derivatives with respect to λ_i and λ_j have the desired sign. To sign the derivative with respect to α, we reason from A1, which says $\lambda_i \ge X\lambda_j$. This implies $\frac{16\theta^2}{X^4\lambda_i^2}(\lambda_i^2 - X^2\lambda_j^2) \ge 0$. Equivalently, $\frac{16}{X^4}\left(\theta^2+\frac{16\alpha^2}{X^4}-\frac{8\alpha\theta\lambda_j}{X\lambda_i}\right) \ge \left(\frac{16\alpha}{X^4}-\frac{4\theta\lambda_j}{X\lambda_i}\right)^2$. This implies $\frac{4}{X^2}\sqrt{\theta^2+\frac{16\alpha^2}{X^4}-\frac{8\alpha\theta\lambda_j}{X\lambda_i}} \ge \frac{16\alpha}{X^4} - \frac{4\theta\lambda_j}{X\lambda_i}$. Thus, we conclude $\frac{\partial s^*}{\partial\alpha} \ge 0$. □

B Additional Details

Appendix B.A summarizes in greater detail the empirical literature on fragmentation that was mentioned in section II. Appendix B.B describes the structure of the Australian market in greater detail. Appendix B.C describes our ASX and Chi-X data in greater detail and explain the steps required to process it. Appendix B.D establishes consistency of our estimation procedure. Appendix B.E describes our implementation of the estimation procedure. Appendix B.F provides details about our selection of the control group used for our analysis of the natural experiment described in section VII.

B.A Empirical Literature on Fragmentation

Paper	Data	Source of Variation
A. Positive association between fragmentation and liquidity		
Branch and Freed (1977)	NYSE and AMEX securities (1974)	cross section
Hamilton (1979)	NYSE equities (1974–1975)	cross section
Neal (1987)	AMEX options (1986–1986)	cross section
Cohen and Conroy (1990)	NYSE equities (1981–1983)	Rule 19c-3
Battalio (1997)	NYSE securities (1988-1990)	Madoff entry
Mayhew (2002)	CBOE options (1986–1997)	panel
Weston (2002)	Nasdaq equities (1998–1999)	panel
Boehmer and Boehmer (2003)	US ETFs (2001)	NYSE entry
De Fontnouvelle et al. (2003)	US options (1999–2000)	cross-listing events
Fink et al. (2006)	Nasdaq equities (1996–2002)	panel
Nguyen et al. (2007)	US options (1993–2002)	panel
Foucault and Menkveld (2008)	Dutch equities (2004–2005)	LSE entry
Chlistalla and Lutat (2011)	French equities (2007)	Chi-X Europe entry
O'Hara and Ye (2011)	US equities (2008)	cross section
Menkveld (2013)	Dutch equities (2007–2008)	Chi-X Europe entry
He et al. (2015)	Global equities (2007–2012)	Chi-X entries (various)
Aitken et al. (2017)	Australian equities (2010–2013)	Chi-X Australia entry
B. Negative association between fragmentation and liquidity		
Bessembinder and Kaufman (1997)	US equities (1994)	Rule 19c-3
Arnold et al. (1999)	US securities (1945–1961)	exchange mergers
Amihud et al. (2003)	Tel-Aviv Stock Exchange warrants (1992–1997)	warrant exercises
Hendershott and Jones (2005)	US ETFs (2002)	Island goes dark
Bennett and Wei (2006)	NYSE & NASDAQ equities (2001–2003)	listing switches
Gajewski and Gresse (2007)	LSE & Euronext Paris equities (2001)	cross section
Nielsson (2009)	European equities (1996–2006)	Euronext mergers
Bernales et al. (2017)	European equities (2008–2009)	Euronext order book consolidation
C. Mixed association between fragmentation and liquidity		
Boneva et al. (2016)	UK equities (2008–2011)	panel
Degryse et al. (2015)	Dutch equities (2007–2009)	panel
Haslag and Ringgenberg (2017)	US securities (1996–2014)	Reg NMS

B.B Industry Background

This appendix describes additional details of the Australian market, including the regulatory environment, the microstructure of ASX and Chi-X, and the off-exchange trading environment.[40]

Market Integrity Rules. The Australian Securities and Investments Commission (ASIC) regulates the Australian market under the Market Integrity Rules (ASIC, 2011). In most respects, these rules are similar to Reg NMS and Reg ATS, their counterparts in the U.S.

[40]Other recent studies of the Australian market include He et al. (2015) and Aitken et al. (2017), both of which we have discussed in the text. In addition, Foley and Putniņš (2016) and Comerton-Forde and Putniņš (2015) also study Australia, although they focus primarily upon dark trading.

However, one notable difference lies in the definition of best execution. For a retail client, ASIC's guidance is that best execution is based on total consideration, typically interpreted as the best average price.[41] (This is in contrast to the U.S., where the order protection rule requires the best marginal price, at least for the top of the book.) For wholesale clients, ASIC's guidance is that best execution can also include factors such as speed. Another notable difference is that payment for order flow is not allowed in Australia.[42] There is also a minimum price improvement rule: trades that take place outside of pre-trade transparent order books must receive price improvement (although there are exceptions for block trades, large portfolio trades, and times outside of trading hours).[43] A consequence of both of these aforementioned facts is that the proportion of retail order flow that executes on exchanges is greater in Australia than in the U.S.

Other relevant aspects of the Market Integrity Rules include a requirement that markets synchronize their clocks to within 20 milliseconds of UTC,[44] a requirement that visible orders have priority over hidden orders at the same price,[45] and a mandated minimum tick size (which, for equities priced above two dollars, is 1 cent).[46]

To cover the costs of market supervision, ASIC imposes fees on market participants. Some of these fees are activity-based, which are calculated on the basis of the market share of trading activity and messaging activity at ASX and Chi-X.

ASX. ASX is the larger and older of Australia's two extant exchanges, having been formed in 1987. It is operated by ASX Limited, which is a publicly traded company. ASX conducts opening and closing auctions. However, the majority of trading takes place in the intervening continuous session. During this session, ASX operates a transparent limit order book called TradeMatch. ASX TradeMatch features pre-trade anonymity for equities, although not for ETFs. In 2011, ASX introduced a second book, PureMatch, which offers fewer functionalities but faster speeds. However, PureMatch failed to attract any significant volume.

Beyond standard limit orders, ASX TradeMatch also offers the following advanced order types: (*i*) iceberg orders, where at least 500 shares must be displayed, (*ii*) undisclosed orders, in which the precise quantity is not disclosed, provided that the value of the order exceeds $0.5 million, and (*iii*) tailor-made combination orders, which can be used for multi-leg transactions.[47]

Chi-X. Chi-X is the smaller and newer of Australia's two exchanges. Like ASX, it is located in Sydney. Chi-X entered the Australian market in 2011. During the sample period, the exchange was operated by a subsidiary of Chi-X Global Holdings LLC, which was privately owned by a consortium of major financial institutions including BofA Merrill Lynch, GETCO

[41] Rule 3.1.1.
[42] Rule 7.5.1.
[43] Rule 4.1.1.
[44] Rule 6.3.1.
[45] Rule 4.1.7.
[46] Rule 6.4.1.
[47] Details on these order types, as well as other aspects of the ASX operating rules are available at ASX (2016).

LLC, Goldman Sachs, Morgan Stanley, Nomura Group, Quantlab Group LP, and UBS. Chi-X Global has made similar entries into other markets, including Europe, Canada, and Japan.[48] Fewer securities are traded on Chi-X than on ASX, and in addition, Chi-X does not perform a listing function. Chi-X offers neither an opening auction nor a closing auction, but just a single limit order book. There is pre-trade anonymity for all securities.

Like ASX, Chi-X offers standard limit orders, as well as iceberg orders and undisclosed orders (but does not offer combination orders). Unlike ASX, Chi-X offers pegged orders, which can reference the bid, ask, or mid price of the national best bid and offer (NBBO). Also unlike ASX, Chi-X allows the placement of completely hidden orders, which interact with visible orders in the book. Chi-X permits broker preferencing for these hidden orders, which allows brokers to cross with their own orders regardless of time priority (yet with regard to price and visibility priority). Chi-X also allows brokers to specify a minimum executable quantity for their orders.[49]

In addition, Chi-X also offers market on close orders, which are fully hidden and trade continuously with each other throughout the day at the ASX closing price (both before and after that price is determined).

Other trading. While the ASX and Chi-X books serve as the main trading venues in Australia, there also exist several other modes of trading, including crossing systems, block trading, and internalization. The largest crossing system is CentrePoint, which is operated by ASX. CentrePoint trades take place at the prevailing TradeMatch mid price, and the venue features full pre-trade anonymity (ASX, 2012a). Other functionalities include (*i*) the ability to specify a minimum executable quantity, (*ii*) broker preferencing, and (*iii*) sweep orders, which allow for simultaneous access of CentrePoint and TradeMatch.

Excluding CentrePoint, twenty other crossing systems, with sixteen separate operators, were active at the beginning of our sample (ASIC, 2016). By the end of our sample, those numbers had fallen to eighteen crossing systems and fifteen operators. The largest of these crossing systems are those operated by Credit Suisse, Goldman Sachs, and Citigroup (ASIC, 2015). These crossing systems account for just 2.4% of total equity market turnover (ASIC, 2015). Additionally, they are not required to provide fair access, and many are accessible by only a small number of traders.

Block trades comprise the majority of off-exchange trading. The minimum size requirement for a block trade is \$0.2 million, \$0.5 million, or \$1 million, depending upon the category of the security in question. Unlike smaller off-exchange trades, which, under the Market Integrity Rules, must receive price improvement relative to the NBBO, block trades can be negotiated at any price.

[48] Chi-X Global has sold all their exchanges since then, although the exchanges continue to operate. In particular, Chi-X Europe was sold to Bats Global Markets in 2011, Chi-X Canada was sold to Nasdaq in 2015, and both Chi-X Australia and Chi-X Japan were sold to private equity firm JC Flowers in 2016.

[49] Details on the operating rules of Chi-X are available at Chi-X (2013).

B.C Data

The raw data are binary encoded files, which constitute the feeds of ASX and Chi-X: "ITCH – Glimpse" (ASX, 2012b) and "Chi-X MD Feed" (Chi-X, 2012), respectively. The outbound feeds of both exchanges are based on NASDAQ's proprietary ITCH protocol. These data are a complete historical record of the information that market participants observe in real-time for a fee. Every trading day is recorded in a separate file, within which messages are recorded chronologically.

These messages are sufficient to construct the lit book at each exchange at any point in time and also to identify all trades that take place in the lit book. Two steps of processing are necessary to obtain that information: message parsing and order book reconstruction. We implement routines to do both using the high-performance computing system *Blacklight* at the Pittsburgh Supercomputing Center, as part of an allocation at XSEDE (Extreme Science and Engineering Discovery Environment).

Message parsing. Every message is binary encoded using MoldUDP64, a networking protocol that allows efficient and scaleable transmission of data. A message is read in as a message block. The first two bytes of the block specify the length of the message, therefore revealing where the message ends and the next begins. The third byte specifies the message type. The interpretation of the remainder of the message depends on the message type. See table 7 for examples of the information embedded in several common message types.

Every second, a timestamp message is broadcast. Other message types include add orders, cancellations, and executions of existing orders. Those messages specify the time, in nanoseconds, relative to the previous timestamp message, as well as any incremental changes to the lit book.[50]

[50] Rather than transmit, for example, the current bid and ask prices, only incremental changes to the book are broadcast. This is to ensure high-performance for latency-sensitive traders.

Table 7: Examples of ASX Message Data Formats

A sample of ASX message specifications (ASX, 2012b). The length of a field is measured in number of bytes.

	length	value
Timestamp Message		
Message Type	1	"T"
Second	4	Numeric
Add Order Message		
Message Type	1	"A"
Timestamp – Nanoseconds	4	Numeric
Order ID	8	Numeric
Order Book ID	4	Numeric
Side (Buy or Sell)	1	Alpha
Order Book Position	4	Numeric
Quantity	8	Numeric
Price	4	Numeric
Order Delete Message		
Message Type	1	"D"
Timestamp – Nanoseconds	4	Numeric
Order ID	8	Numeric
Order Book ID	4	Numeric
Side (Buy or Sell)	1	Alpha
Side (Buy or Sell)	1	Alpha

Order book reconstruction. To reconstruct the limit order books, we follow the detailed instructions in ASX (2012b, section 2.9) and Chi-X (2012, section 5). We outline the steps below. Each message conveys only an incremental change. Therefore, reconstructing the lit order book for a given day and security requires re-running the message broadcast in chronological order beginning at market open and replicating the matching process used by the exchange. When an add order arrives, it is added to the book at the limit price that it specifies. In case of a cancellation, the active order in question is removed. Finally, in the event of an execution, the affected order is removed or its quantity is adjusted. The time series of inside quotes can then be computed by reading off the bid and ask that prevail after the processing of each message.

Lit book volume. The exchange feeds distinguish between two types of trades. First are on-exchange trades in which the passive order had been visible (either a fully visible order or the visible portion of an iceberg order). Our analysis focuses solely on these trades, which we call *lit book volume*.[51] As discussed in section V.B, the remainder includes (*i*) trading in the ASX opening and closing crosses, (*ii*) off-exchange trading (e.g., trading in crossing systems, block trades, and internalization), and (*iii*) on-exchange trades in which the passive order

[51] In the ASX feed, these are the trades associated with "E" or "C" messages (ASX, 2012b, section 2.6.2), and in the Chi-X feed, these are the trades associated with "E" messages (Chi-X, 2012, section 5.3).

had not been visible (either a fully hidden order, the hidden portion of an iceberg order, or an undisclosed order).

B.D Consistency

This appendix establishes consistency of the nonlinear least squares (NLS) estimation procedure described in section VI.A. While we assume that, conditional on the quotes, both $\varepsilon_{x,t}^{\text{buy}}$ and $\varepsilon_{x,t}^{\text{sell}}$ have mean zero, it would not be correct to make the same assumption for $\varepsilon_{x,t}^{\text{spread}}$. Therefore, consistency of our estimation procedure does not follow immediately from the standard arguments that typically establish consistency of NLS. Nevertheless, consistency is restored by two special features of the model. First, λ_j is excluded from equations (8) and (9). Second, we have the following property related to equation (10): for $(\hat{\alpha}, \hat{\theta}, \hat{\lambda}_i)$ in a neighborhood of $(\alpha, \theta, \lambda_i)$, $\hat{\theta} + \hat{\alpha} - \sqrt{\hat{\theta}^2 + \hat{\alpha}^2 - 4\hat{\alpha}\hat{\theta}\hat{\lambda}_j/\hat{\lambda}_i}$ is an injective function of $\hat{\lambda}_j$ in a neighborhood of λ_j, which contains a neighborhood of $\theta + \alpha - \sqrt{\theta^2 + \alpha^2 - 4\alpha\theta\lambda_j/\lambda_i}$ in its range.

Proposition 5. *Assuming that the data generation process is as described in section VI.A, where the parameters satisfy A1, A2 and A3, the NLS estimation procedure described in section VI.A is consistent for those parameters.*

Proof. Define

$$Q_T(\alpha, \theta, \lambda_i, \lambda_j) = \frac{1}{T}\sum_{t=1}^{T}\sum_{x\in\{\text{ASX,Chi-X}\}} \left(\varepsilon_{x,t}^{\text{buy}}\right)^2 + \left(\varepsilon_{x,t}^{\text{sell}}\right)^2 + \left(\varepsilon_{x,t}^{\text{spread}}\right)^2$$

and

$$\tilde{Q}_T(\alpha, \theta, \lambda_i) = \frac{1}{T}\sum_{t=1}^{T}\sum_{x\in\{\text{ASX,Chi-X}\}} \left(\varepsilon_{x,t}^{\text{buy}}\right)^2 + \left(\varepsilon_{x,t}^{\text{sell}}\right)^2,$$

where $\varepsilon_{x,t}^{\text{buy}}$, $\varepsilon_{x,t}^{\text{sell}}$, and $\varepsilon_{x,t}^{\text{spread}}$ are as defined implicitly by equations (8), (9), and (10), respectively. The NLS estimates are then a selection

$$(\hat{\alpha}, \hat{\theta}, \hat{\lambda}_i, \hat{\lambda}_j) \in \underset{\alpha,\theta,\lambda_i,\lambda_j}{\arg\min}\, Q_T(\alpha, \theta, \lambda_i, \lambda_j).$$

Define also

$$\bar{s} = \frac{1}{2T}\sum_{t=1}^{T}\sum_{x\in\{\text{ASX,Chi-X}\}} s_{x,t}$$

Finally, define

$$\tilde{Q}_T^* = \left\{ (\tilde{\alpha}, \tilde{\theta}, \tilde{\lambda}_i, \tilde{\lambda}_j) \;\middle|\; \begin{array}{l} (\tilde{\alpha}, \tilde{\theta}, \tilde{\lambda}_i) \in \arg\min_{\alpha,\theta,\lambda_i} \tilde{Q}_T(\alpha, \theta, \lambda_i), \\ \text{and } \tilde{\lambda}_j = \frac{\tilde{\lambda}_i}{4\tilde{\alpha}\tilde{\theta}}\left(2\tilde{\alpha}\bar{s} + 2\tilde{\theta}\bar{s} - \bar{s}^2 - 2\tilde{\alpha}\tilde{\theta}\right) \end{array} \right\}.$$

With these definitions in hand, we complete the proof by establishing two claims.

Claim 1: Any selection $(\tilde{\alpha}, \tilde{\theta}, \tilde{\lambda}_i, \tilde{\lambda}_j) \in \tilde{Q}_T^*$ is consistent for $(\alpha, \theta, \lambda_i, \lambda_j)$.

Proof of claim: By the assumptions imposed upon $\varepsilon_{x,t}^{\text{buy}}$ and $\varepsilon_{x,t}^{\text{sell}}$, the usual arguments for consistency of NLS establish that

$$(\tilde{\alpha}, \tilde{\theta}, \tilde{\lambda}_i) \xrightarrow{p} (\alpha, \theta, \lambda_i).$$

Standard arguments also imply

$$\bar{s} \xrightarrow{p} \theta + \alpha - \sqrt{\theta^2 + \alpha^2 - \frac{4\alpha\theta\lambda_j}{\lambda_i}}.$$

Thus, the continuous mapping theorem yields

$$\tilde{\lambda}_j \xrightarrow{p} \lambda_j.$$

Claim 2: The NLS estimates $(\hat{\alpha}, \hat{\theta}, \hat{\lambda}_i, \hat{\lambda}_j)$ converge almost surely, and hence in probability, to $\tilde{Q}_T^*$.

Proof of claim: First note that

$$\begin{aligned} Q_T(\hat{\alpha}, \hat{\theta}, \hat{\lambda}_i, \hat{\lambda}_j) &= \tilde{Q}_T(\hat{\alpha}, \hat{\theta}, \hat{\lambda}_i) + \frac{1}{T}\sum_{t=1}^{T} \sum_{x \in \{\text{ASX,Chi-X}\}} \left(s_{x,t} - \hat{\theta} - \hat{\alpha} + \sqrt{\hat{\theta}^2 + \hat{\alpha}^2 - \frac{4\hat{\alpha}\hat{\theta}\hat{\lambda}_j}{\hat{\lambda}_i}} \right)^2 \\ &\geq \min_{\alpha,\theta,\lambda_i} \tilde{Q}_T(\alpha, \theta, \lambda_i) + \frac{1}{T}\sum_{t=1}^{T} \sum_{x \in \{\text{ASX,Chi-X}\}} \left(s_{x,t} - \bar{s}\right)^2. \end{aligned} \tag{13}$$

A parameter vector $(\hat{\alpha}, \hat{\theta}, \hat{\lambda}_i, \hat{\lambda}_j)$ achieves the lower bound (13) only if both of the following conditions hold:

$$(\hat{\alpha}, \hat{\theta}, \hat{\lambda}_i) \in \arg\min_{\alpha,\theta,\lambda_i} \tilde{Q}_T(\alpha, \theta, \lambda_i)$$

$$\hat{\theta} + \hat{\alpha} - \sqrt{\hat{\theta}^2 + \hat{\alpha}^2 - \frac{4\hat{\alpha}\hat{\theta}\hat{\lambda}_j}{\hat{\lambda}_i}} = \bar{s}$$

Note that the second of these conditions holds only if

$$\hat{\lambda}_j = \frac{\hat{\lambda}_i}{4\hat{\alpha}\hat{\theta}} \left(2\hat{\alpha}\bar{s} + 2\hat{\theta}\bar{s} - \bar{s}^2 - 2\hat{\alpha}\hat{\theta}\right).$$

Thus, the lower bound (13) is achieved only if $(\hat{\alpha}, \hat{\theta}, \hat{\lambda}_i, \hat{\lambda}_j) \in \tilde{Q}_T^*$.

Let $\delta > 0$ and take a selection $(\tilde{\alpha}, \tilde{\theta}, \tilde{\lambda}_i, \tilde{\lambda}_j) \in \tilde{Q}_T^*$. We have seen that $\bar{s} \xrightarrow{p} \theta + \alpha - \sqrt{\theta^2 + \alpha^2 - 4\alpha\theta\lambda_j/\lambda_i}$ and that $\tilde{\theta} + \tilde{\alpha} \xrightarrow{p} \theta + \alpha$. Thus, $\tilde{\theta} + \tilde{\alpha} - \bar{s} \xrightarrow{p} \sqrt{\theta^2 + \alpha^2 - 4\alpha\theta\lambda_j/\lambda_i}$. By lemma 1(iv), this limit is strictly positive. There thus exists some T' such that if $T \geq T'$, then $\Pr(\bar{s} \leq \tilde{\theta} + \tilde{\alpha}) > 1 - \delta$. In such circumstances,

$$\tilde{\theta} + \tilde{\alpha} - \sqrt{\tilde{\theta}^2 + \tilde{\alpha}^2 - \frac{4\tilde{\alpha}\tilde{\theta}\tilde{\lambda}_j}{\tilde{\lambda}_i}} = \bar{s}.$$

Since we also have

$$(\tilde{\alpha}, \tilde{\theta}, \tilde{\lambda}_i) \in \underset{\alpha,\theta,\lambda_i}{\arg\min}\, \tilde{Q}_T(\alpha, \theta, \lambda_i),$$

$(\tilde{\alpha}, \tilde{\theta}, \tilde{\lambda}_i, \tilde{\lambda}_j)$ achieves the lower bound (13). To summarize: in such circumstances, the lower bound (13) is achievable, and so any selection $(\hat{\alpha}, \hat{\theta}, \hat{\lambda}_i, \hat{\lambda}_j) \in \arg\min_{\alpha,\theta,\lambda_i,\lambda_j} Q_T(\alpha, \theta, \lambda_i, \lambda_j)$ must achieve it and must therefore be contained in $\tilde{Q}_T^*$. □

B.E Estimation

Estimation procedure. Estimation is conducted by minimizing the NLS objective function (11), as described in section VI.A. This minimization was performed using SNOPT (Gill, Wong, Murray and Saunders, 2015). To evade the possibility of identifying a local minimum that is not the global minimum, we repeat the optimization for 200 different randomly chosen starting values. During estimation we impose non-negativity constraints on the parameters.

Standard errors. We use a bootstrap procedure to compute standard errors. To allow for temporal dependence, we use a block bootstrap. The asymptotically optimal block length grows with the sample size T at a rate proportional to $T^{1/3}$ (Hall, Horowitz and Jing, 1995). In our case, $T = 1,584,000$, which gives $T^{1/3} \approx 117$. So as to avoid blocks that span days, which would not correctly capture the dependence in the data, we use non-overlapping blocks and round to a block length of 120 seconds. Thus, each of the 80 trading days in the sample is divided into 165 blocks. For each bootstrap replication we draw 165×80 blocks with replacement. Standard errors are computed based on 200 bootstrap replications.

Discussion of estimation approach. GMM is a popular class of estimators, which has been used to estimate other models of limit order book trading (e.g. Sandås, 2001; Biais, Bisière and Spatt, 2010). Our estimation procedure is also encompassed by the GMM framework, with the sample moment conditions being the first order conditions of equation (11) with respect to the parameters.[52]

In addition, others have used MLE to estimate models of limit order book trading (e.g. Glosten and Harris, 1988; Easley, Kiefer, O'Hara and Paperman, 1996; Easley, Engle, O'Hara and Wu, 2008). In principle, we could have done the same, but this would require parametrizing the error terms in equations (8), (9), and (10). Since our sample is large, it seems unlikely that the efficiency gains brought by MLE would be sufficient to justify additional distribu-

[52] Establishing that uses the fact that the error terms in equations (8), (9), and (10) enter additively (Cameron and Trivedi, 2005, p. 168).

tional assumptions.

B.F Control Group Selection

In this appendix we provide details about the selection of the control group that we have used in the analysis of the Chi-X shutdown in section VII, where the results of that analysis are reported in column (5) of table 6. The control group consists of eight securities that are similar to STW in that they are also ETFs with exposure to Australian equities. However, unlike STW, they were not traded on Chi-X in June 2014.

To construct the control group, we start from a list of all 17 ETFs traded on ASX with exposure to Australian equity securities as of February 2017 (ASX, 2017). One of these ETFs is STW, the focus of our analysis. Six of these ETFs were admitted only after June 2014, and are therefore discarded. Another two were also traded on Chi-X as of June 2014, so they are discarded as well, which leaves us with the remaining eight ETFs that constitute the control group for our analysis. Table 8 summarizes them.

Table 8: Australian Equity ETFs, June 2014

The control group is selected from the set of ETFs with with exposure to Australian equities provided by ASX (2017). Such an ETF is selected into the control group if (*i*) it existed in June 2014; and (*ii*) it was traded on ASX but not on Chi-X as of May 30, 2014. An ETF's admission date refers to when it was first admitted for trading at ASX.

Code	Benchmark	Admission Date
A. Control Group		
IOZ	S&P/ASX 200	Dec 2010
ISO	S&P/ASX Small Ordinaries	Dec 2010
MVW	MVIS Australia Equal Weight Index	Mar 2014
QOZ	FTSE RAFI Australia 200	Jul 2013
SSO	S&P/ASX Small Ordinaries	Apr 2011
VAS	S&P/ASX 300	May 2009
VLC	MSCI Large Cap Index	May 2011
VSO	MSCI Small Cap Index	May 2011
B. Treatment Group		
STW	S&P/ASX 200	Aug 2001

C Additional Counterfactuals

Appendix C.A investigates the consequences of replacing the limit order book with either of two counterfactual trading mechanisms: frequent batch auctions or non-cancellation delays. Appendix C.B investigates the counterfactual in which an order protection rule is applied to the Australian market.

C.A Alternative Trading Mechanisms

A current debate among policy makers, industry participants, and researchers concerns whether alternative trading mechanisms can improve upon the prevailing limit order book. In this section, we consider the counterfactual of the model under two of those alternatives: frequent batch auctions and non-cancellation delays.

The frequent batch auction mechanism would replace the limit order book with sealed-bid, uniform price double auctions conducted at discrete intervals. Frequent batch auctions are the focus of Budish et al. (2015), who show that they improve upon the limit order book by eliminating stale-quote sniping on the basis of public news, and the same remains true in our extension of their framework. The intuition is as follows. Frequent batch auctions delay the processing of all orders received during a batch interval until the end of that interval, which ensures that orders submitted at the same time are processed together. This effectively allows liquidity providers to update their stale quotes before they can be sniped, thereby eliminating this source of adverse selection.

The non-cancellation delay mechanism would modify the limit order book by adding a small, possibly random, delay to all orders except cancellations. It is considered by Baldauf and Mollner (forthcoming), who show that it also eliminates what they refer to as "aggressive-side order anticipation." But, more relevant to the model considered in this paper, it also eliminates stale-quote sniping on the basis of public news. The intuition is as follows. When the liquidity provider submits an order to cancel a mispriced quote at the same time that a sniper submits an order to trade against that mispriced quote, either could be processed first under the limit order book mechanism. In contrast, under non-cancellation delays, the cancellation is guaranteed to be processed first.

In this model, stale-quote sniping is the only source of adverse selection. Thus, the consequences of eliminating it are mathematically equivalent to what would transpire if there were never any jumps in the fundamental value of the security (i.e., $\lambda_j = 0$). While this can be established formally, we omit the derivation in the interest of brevity. Given this, immediate corollaries of propositions 1 and 2 are the following characterizations of the spread that prevails under either frequent batch auctions or non-cancellation delays.

Corollary 6. *With a single exchange ($X = 1$) that uses either frequent batch auctions or non-cancellation delays, there exists a SPNE with spread*

$$s^*_{FBA} = s^*_{ND} = \theta.$$

Corollary 7. *With multiple exchanges ($X \geq 2$) that use either frequent batch auctions or non-cancellation delays, there exists a SPNE with spread*

$$s^*_{FBA} = s^*_{ND} = \theta + \frac{2\alpha}{X} - \sqrt{\theta^2 + \frac{4\alpha^2}{X^2}}.$$

In addition, an immediate corollary of proposition 4 is that frequent batch auctions and non-cancellation delays result in a spread that is guaranteed to be smaller than that which prevails under the limit order book. The intuition is that, by eliminating stale-quote sniping, these alternative trading mechanisms eliminate the portion of the spread stemming from adverse selection, leaving only the portion stemming from the market power of exchanges.

Corollary 8. $s^*_{FBA} = s^*_{ND} \leq s^*$.

Theory therefore dictates that either frequent batch auctions or non-cancellation delays would improve outcomes by reducing transaction costs. What is more, our empirical approach can quantify this reduction. Evaluating the expressions from corollaries 6 and 7 at the estimated parameters, we find that in the counterfactual of frequent batch auctions or non-cancellation delays, the duopoly spread is 52.4% lower relative to the spread in the prevailing limit order book duopoly (*cf.* column 2 in table 9).

Moreover, since frequent batch auctions and non-cancellation delays eliminate all adverse selection from the model, they also shut down the exposure channel. Thus, fragmentation unambiguously lowers spreads under the regimes of frequent batch auctions and non-cancellation delays. This reverses the ranking of the duopoly and monopoly spreads relative to what prevailed under the limit order book. Under these alternative mechanisms, moving from duopoly to monopoly *raises* spreads by 11.5%, from 1.37 to 1.53¢. Furthermore, a triopoly is not only feasible, but also would feature spreads that are still lower.

Table 9: Cum-Fee Spreads Under Various Counterfactuals (Cents)

The counterfactual spreads are based on the parameter estimates in table 4. Rows refer to trading mechanisms, and columns refer to the number of exchanges. At the parameter estimates, an equilibrium with three exchanges actively operating limit order books is inconsistent with A2 and A3.

	number of exchanges		
	1	2	3
limit order book	2.22	2.88	
frequent batch auctions/non-cancellation delays	1.53	1.37	1.19

C.B Order Protection Rule

One of the stipulations of Reg NMS in the U.S. is the order protection rule (also known as Rule 611, or the trade-through rule). It requires trading venues to maintain and enforce procedures that limit the possibility of a trade occurring on that venue when a better price is available elsewhere. In contrast, the legal framework governing trading in Australia does not currently incorporate an order protection rule. In this appendix, we use the estimated model to shed some light on the counterfactual in which Australia were to adopt such a rule.

The most natural way to interpret an order protection rule within the model would be as a reduction in α, the magnitude of the frictions that prevent investors from filling their orders at the best price. Proposition 4 implies that such a reduction in α would lead to a reduction of the spread. The remainder of this appendix is dedicated to quantifying the extent of this reduction.

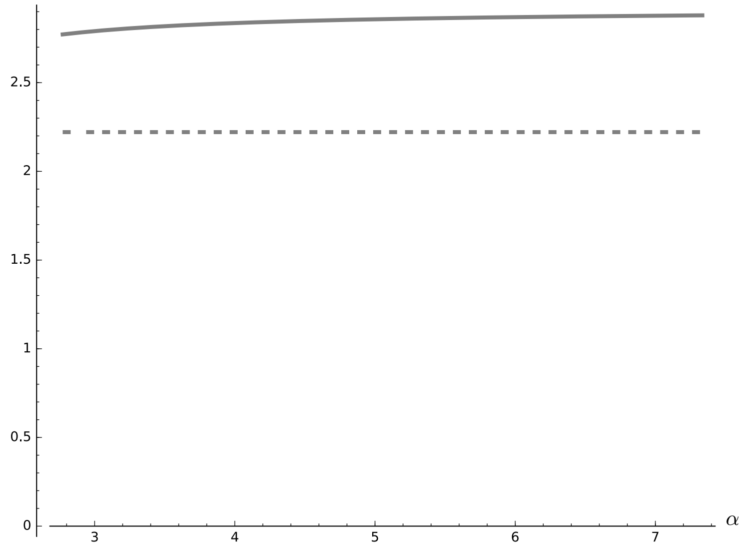
It is unclear what the precise magnitude of the reduction in α would be. In particular, we likely would not expect a complete eradication of market frictions (i.e., a full reduction of α to zero). Rather, some residual frictions would likely remain, due, for example, to (*i*) difficulties with monitoring prices in real time, (*ii*) imperfect enforcement of the rule, or (*iii*) discrete prices.

Moreover, the value of α cannot be reduced arbitrarily without violating A3, evaluated in the case of a duopoly. If the effect of the rule were to reduce α below this threshold, then competition among exchanges would be intensified to such an extent that it would be impossible for both ASX and Chi-X to operate profitably, and we might expect a reduction in the number of exchanges in the long run. In such cases, the relevant counterfactual would be the monopoly case. On the other hand, if α remains above this threshold, then exchange profitability would be reduced but might not be completely eliminated. The precise value of this threshold depends upon σ. While our procedure does not produce an estimate of σ, a lower bound for it is given by A2. Evaluating at that lower bound, we obtain a corresponding lower bound for the threshold: $\bar{\alpha} = 2.772$ is the smallest value of α that is consistent with the model assumptions at the parameter estimates, such that a duopoly remains feasible.

Figure 3 illustrates the counterfactual analysis. The solid line represents the spread that would prevail in a duopoly for values of α below that estimated in the data, yet above $\bar{\alpha}$. This is the relevant counterfactual if A3 remains satisfied given the new value of α and given the value of σ. The dashed line depicts the monopoly spread. It is the relevant counterfactual if A3 is violated given the new value of α and given the value of σ. Interestingly, this analysis suggests that the effects of an order protection rule would be fairly minimal, except for the case in which the rule induces the exit of an exchange. The reason for this is that the estimated model indicates that existing competition between ASX and Chi-X is such that their profit margins are already fairly thin. There is therefore not much scope for intensifying this competition without eliminating exchange profitability altogether.

Figure 3: Counterfactual Spreads with Order Protection

The figure plots the counterfactual spreads for monopoly (dashed line) and duopoly (solid line) for values of α below the estimate of the parameter (*cf.* table 4), yet large enough to be consistent with the estimates of the other parameters and the assumptions of the model (*viz.* A2 and A3).



D Robustness of Estimates

In appendix D.A, we demonstrate that the results of our estimation procedure are robust to alternative choices of the cutoff that is used to distinguish between isolated and clustered trades. In appendix D.B, we demonstrate that our results are also robust to using only buys or only sells as the basis for estimation, and we also show that we fail to reject an overidentifying restriction. Finally, in appendix D.C, we also demonstrate robustness with respect to the timeframe of the sample.

D.A Robustness to Classification Error

For the empirical analysis in the main text, we use a one second cutoff for distinguishing between isolated and clustered trades: a lit book trade is classified as isolated if no other such trade occurs in the same direction within one second on either exchange, and it is classified as clustered otherwise. This appendix demonstrates that the main results are robust to changes in this cutoff.

Column (3) of table 10 contains the baseline results reported in the main text. To construct the remaining columns, we repeat our estimation procedure for four alternative choices of this cutoff, ranging from 0.1 seconds to 5 seconds. The table reveals that the precise definition of what separates an isolated trade from a clustered trade—at least within this range—has very little impact on the parameter estimates and, consequently, does not change the results qualitatively. Over these different robustness checks, the counterfactual monopoly spread is never smaller than 2.19¢ and never larger than 2.28¢.

Table 10: Estimates for Different Definitions of Isolated/Clustered Trades

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Point estimates are computed to minimize the objective (11). Standard errors are based on 200 block bootstrap replications. Columns refer to five different definitions of what constitutes an isolated trade. Panel B shows the corresponding counterfactual monopoly and duopoly spreads (calculated from propositions 1 and 2).

	value of cutoff (seconds)				
	0.1	0.5	1	2	5
A. parameter estimates					
α	9.64338***	4.92955***	7.33504***	12.10799***	5.48684***
	(1.45972)	(0.71493)	(1.06129)	(1.77839)	(0.76835)
θ	1.6213***	1.57209***	1.52817***	1.54993***	1.48682***
	(0.16318)	(0.17743)	(0.15127)	(0.14898)	(0.16159)
λ_i	0.00184***	0.00176***	0.00172***	0.00166***	0.00164***
	(0.00024)	(0.00024)	(0.00022)	(0.00022)	(0.00023)
λ_j	0.00074***	0.00077***	0.00078***	0.00072***	0.00078***
	(0.00012)	(0.00012)	(0.00012)	(0.00012)	(0.00011)
B. counterfactual spreads					
monopoly	2.27722***	2.26421***	2.22093***	2.22753***	2.19527***
	(0.1065)	(0.14064)	(0.10647)	(0.09287)	(0.12471)
duopoly	2.87885***	2.87885***	2.87885***	2.87885***	2.87885***
	(0.00476)	(0.00457)	(0.0048)	(0.00497)	(0.0046)

D.B Robustness: Side of the Book

In the main text, we report parameter estimates based on an estimation procedure that leverages both buys and sells. In this appendix, we first demonstrate that we obtain similar results from estimation procedures that leverage only buys or only sells. We then use this framework as the basis for testing an overidentifying restriction implied by the model.

Column (3) of table 11 contains the baseline results reported in the main text. Columns (1) and (2) report the results obtained from analogous estimation procedures that are based on only buys or only sells, respectively. To describe those estimation procedures in more detail, consider the following estimating equations:

$$buy_{x,t} = \frac{\lambda_i^B}{2}\left[\frac{1}{2} + \frac{a_{-x,t} - a_{x,t}}{\alpha^B}\right]_0^1 \left[1 - \frac{a_{x,t} - v_t}{\theta^B}\right]_0^1 + \varepsilon_{x,t}^{\text{buy}} \tag{14}$$

$$sell_{x,t} = \frac{\lambda_i^S}{2}\left[\frac{1}{2} + \frac{b_{x,t} - b_{-x,t}}{\alpha^S}\right]_0^1 \left[1 - \frac{v_t - b_{x,t}}{\theta^S}\right]_0^1 + \varepsilon_{x,t}^{\text{sell}} \tag{15}$$

$$s_{x,t} = \theta^B + \alpha^B - \sqrt{(\theta^B)^2 + (\alpha^B)^2 - \frac{4\alpha^B\theta^B\lambda_j^B}{\lambda_i^B}} + \varepsilon_{x,t}^{\text{spread,buy}}$$

$$s_{x,t} = \theta^S + \alpha^S - \sqrt{(\theta^S)^2 + (\alpha^S)^2 - \frac{4\alpha^S\theta^S\lambda_j^S}{\lambda_i^S}} + \varepsilon_{x,t}^{\text{spread,sell}}$$

As before, t indexes the seconds in the sample and x indexes the exchanges $\{\text{ASX}, \text{Chi-X}\}$. And as before, we proxy for v_t with the average mid price $(b_{\text{ASX},t} + b_{\text{Chi-X},t} + a_{\text{ASX},t} + a_{\text{Chi-X},t})/4$ in both (14) and (15).

The estimation procedure for buys is to minimize the objective

$$Q_T^B(\alpha^B, \theta^B, \lambda_i^B, \lambda_j^B) = \frac{1}{T}\sum_{t=1}^{T}\sum_{x\in\{\text{ASX,Chi-X}\}} \left(\varepsilon_{x,t}^{\text{buy}}\right)^2 + \frac{1}{2}\left(\varepsilon_{x,t}^{\text{spread,buy}}\right)^2$$

Likewise, the estimation procedure for sells is to minimize the objective

$$Q_T^S(\alpha^S, \theta^S, \lambda_i^S, \lambda_j^S) = \frac{1}{T}\sum_{t=1}^{T}\sum_{x\in\{\text{ASX,Chi-X}\}} \left(\varepsilon_{x,t}^{\text{sell}}\right)^2 + \frac{1}{2}\left(\varepsilon_{x,t}^{\text{spread,sell}}\right)^2$$

Comparing across the columns of table 11, the point estimates do differ somewhat depending on the estimation procedure. However, the counterfactual monopoly spreads are remarkably similar.

Note that we can perform the buy and sell estimation procedures jointly by minimizing the objective

$$Q_T^J(\alpha^B, \theta^B, \lambda_i^B, \lambda_j^B, \alpha^S, \theta^S, \lambda_i^S, \lambda_j^S) = \frac{1}{T}\sum_{t=1}^{T}\sum_{x\in\{\text{ASX,Chi-X}\}}\sum_{i\in\{\text{buy,sell}\}} \left(\varepsilon_{x,t}^{i}\right)^2 + \frac{1}{2}\left(\varepsilon_{x,t}^{\text{spread},i}\right)^2$$

Doing so allows us to test the restriction that the parameters in the buy and sell equations are equal: $\alpha^B = \alpha^S$, $\theta^B = \theta^S$, $\lambda_i^B = \lambda_i^S$, $\lambda_j^B = \lambda_j^S$. To do so, we perform a standard Wald test of this joint restriction using the bootstrapped variance-covariance matrix. We compute a test statistic of 6.75, based on which we conclude that we fail to reject the Null of parameter equality, even at the ten percent level.[53]

Finally, note that the baseline estimation is equivalent to performing the above joint

[53]The critical values of the χ^2 distribution with four degrees of freedom are 13.28, 9.49, and 7.78, for a 1%, 5%, or 10% test, respectively.

estimation with the restrictions in place. As a result, the aforementioned test can also be interpreted as a test of an overidentifying restriction for our estimation procedure.

Table 11: Estimates Based on Buys and Sells

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Point estimates are computed to minimize the objective (11). Standard errors are based on 200 block bootstrap replications. Columns refer to different objective functions. Column (3) is based on minimizing the objective Q_T, as described in section VI.A. Columns (1) and (2) are based on minimizing the objectives Q_T^B and Q_T^S, respectively, as described in this appendix. Panel B shows the corresponding counterfactual monopoly and duopoly spreads (calculated from propositions 1 and 2).

	Buy	Sell	Joint
A. parameter estimates			
α	8.73269***	4.92870***	7.33504***
	(1.25134)	(0.70999)	(1.06129)
θ	1.52424***	1.56523***	1.52817***
	(0.15689)	(0.18383)	(0.15127)
λ_i	0.00162***	0.00173***	0.00172***
	(0.00023)	(0.00022)	(0.00022)
λ_j	0.00073***	0.00077***	0.00078***
	(0.00011)	(0.00013)	(0.00012)
B. counterfactual spreads			
monopoly	2.21552***	2.25878***	2.22093***
	(0.10531)	(0.14853)	(0.10647)
duopoly	2.87885***	2.87885***	2.87885***
	(0.00478)	(0.00480)	(0.00480)

D.C Robustness: Timeframe

In table 12 we investigate whether the parameter estimates are constant over time. Column (5) of the table contains the baseline results reported in the main text. To construct the remaining columns, we repeat our estimation procedure for subsamples of trading days pertaining to February, March, April, and May of 2014. Although some of the point estimates differ, the monopoly spread, which is our key counterfactual, remains quite stable across subsamples.

Table 12: Estimates for Different Months

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Point estimates are computed to minimize the objective (11). Standard errors are based on 200 block bootstrap replications. Columns refer to different samples. Panel B shows the corresponding counterfactual monopoly and duopoly spreads (calculated from propositions 1 and 2).

	February	March	April	May	Full
A. parameter estimates					
α	7.33507***	9.64332***	4.92913***	12.108***	7.33504***
	(1.05744)	(1.429)	(0.69578)	(1.73488)	(1.06129)
θ	1.52989***	1.62702***	1.56615***	1.5516***	1.52817***
	(0.15613)	(0.16883)	(0.16855)	(0.16920)	(0.15127)
λ_i	0.00172***	0.00184***	0.00177***	0.00166***	0.00172***
	(0.00023)	(0.00025)	(0.00023)	(0.00023)	(0.00022)
λ_j	0.00078***	0.00075***	0.00076***	0.00073***	0.00078***
	(0.00013)	(0.00013)	(0.00012)	(0.00012)	(0.00012)
B. counterfactual spreads					
monopoly	2.22558***	2.29256***	2.24185***	2.23249***	2.22092***
	(0.11048)	(0.11138)	(0.13515)	(0.10489)	(0.10647)
duopoly	2.88732***	2.906***	2.83104***	2.88767***	2.87884***
	(0.01248)	(0.00932)	(0.00911)	(0.01032)	(0.00480)

E Additional Evidence

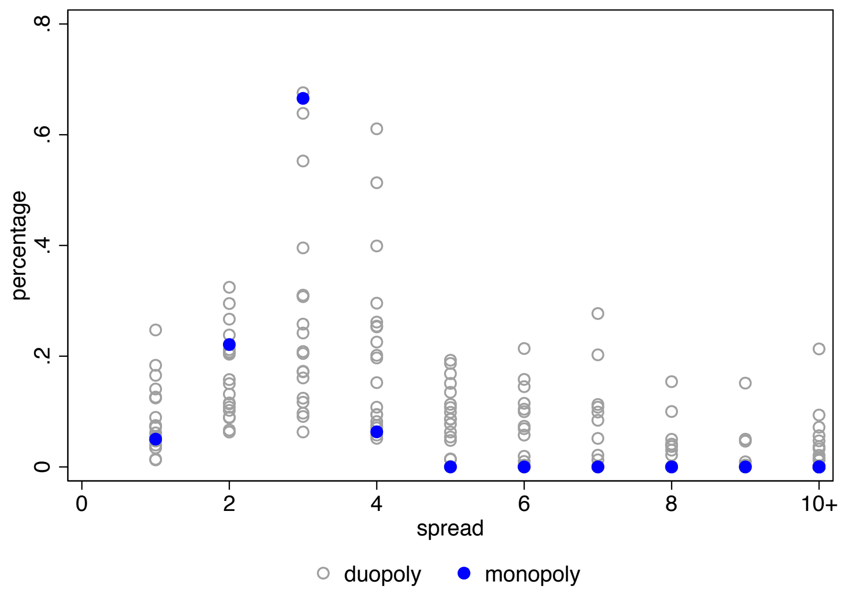
Appendix E.A provides a graph to compare the distribution of the ASX spread on the day of the Chi-X shutdown to the corresponding distribution on the surrounding days. Appendix E.B demonstrates that the results we obtain in our analysis of the Chi-X shutdown (*cf.* section VII.B) survive even if we control for traded volume in various ways. Appendix E.C presents reduced form evidence of the own-price and cross-price elasticities, which are at the heart of the demand system for investors that is postulated by the model. Finally, appendix E.D demonstrates that clustered trades predict subsequent price movements better than isolated trades, which supports our use of isolated trades as a proxy for trades that are precipitated by liquidity-motivated investors and clustered trades as a proxy for trades that are precipitated by information-motived snipers.

E.A Natural Experiment: Distribution of ASX Spread

Figure 4 displays the probability mass function of the empirical distribution of the quoted spread on ASX for each of the trading days of June 2014. In the figure, the day of the Chi-X shutdown is labeled "monopoly," and remaining days are labeled "duopoly." Consistent with the analysis of section VII.B, the figure illustrates that quoted spreads are on average lower on the monopoly day than on the surrounding duopoly days. In particular, the right tail of the spread distribution is much thinner on the monopoly day.

Figure 4: Distribution of ASX Quoted Spread of STW, June 2014

Probability mass functions of the empirical distribution of quoted bid-ask spreads (cents) of STW on ASX for each of the 20 trading days of June 2014. For each day, the sample comprises all seconds between 11:08 and 16:00.



E.B Natural Experiment: Controlling for Volume

In terms of total volume, the monopoly day is in the bottom quartile of the trading days in June 2014. A potential concern is therefore that our findings are driven not by the Chi-X shutdown but by alternative factors that are correlated with low volume. Table 13 addresses this concern in two ways. First, we show that our results survive even if we restrict our analysis to samples of low-volume days. Columns (1) and (2) replicate columns (1) and (5) of table 6 for days in the bottom quartile with respect to total volume. Columns (3) and (4) do the same for days in the bottom half. Second, we show that our results survive even if volume is added as a control variable. Column (5) adds volume as a control but otherwise replicates column (1) of table 6.

Table 13: ASX spreads, Australian Equity ETFs, June 2014
(Controlling for Volume)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the cum-fee spread in cents prevailing on ASX at the beginning of the second. An observation is a pair: a second between 11:08 and 16:00 in June 2014 and a security traded on ASX. STW × Monopoly is an indicator for June 16, 2014 and STW. Volume is total STW volume for Australia, as obtained from Bloomberg. Samples for the respective columns are: (1) trading days on which total STW volume was in the bottom quartile of June 2014 and STW, (2) trading days on which total STW volume was in the bottom quartile of June 2014 and the securities STW, IOZ, ISO, MVW, QOZ, SSO, VAS, VLC, and VSO, (3) trading days on which total STW volume was in the bottom half of June 2014 and STW, (4) trading days on which total STW volume was in the bottom half of June 2014 and the securities STW, IOZ, ISO, MVW, QOZ, SSO, VAS, VLC, and VSO, (5) all trading days in June 2014 and STW. Coefficients are estimated by ordinary least squares. Standard errors are clustered by 120 second blocks on each trading day. STW Duopoly Mean is the average of the dependent variable for STW, non-monopoly observations. Change (Percent) is the estimate relative to the STW duopoly mean.

	(1)	(2)	(3)	(4)	(5)
	[Volume: bottom quartile]		[Volume: bottom half]		Full sample
STW × Monopoly	-0.482***	-0.376***	-0.695***	-0.561***	-0.824***
	(0.0642)	(0.0725)	(0.0607)	(0.0671)	(0.0601)
Volume					0.00410***
					(0.000487)
STW Duopoly Mean	3.378	3.378	3.590	3.590	4.023
Change (Percent)	-14.27	-11.12	-19.35	-15.63	-20.48
Day × Hour Fixed Effects	NO	YES	NO	YES	NO
Security Fixed Effects	YES	YES	YES	YES	YES
Control Group	NO	YES	NO	YES	NO
Observations	87600	788400	175200	1576800	350400

E.C Reduced Form Evidence of Own- and Cross-Price Elasticities

The driving force of the model is a demand system that governs the exchange choice of investors. In particular, the model predicts an own-price elasticity (i.e., fewer investors trade at an exchange when its prices become less favorable) as well as a cross-price elasticity (i.e., more investors trade at an exchange when prices at other exchanges become less favorable). In this appendix, we present reduced form evidence for the existence of these elasticities in the data.

Table 14 displays the coefficients of regressions that explain variation in the occurrence of isolated buys and sells at ASX and Chi-X in terms of the prices prevailing at the two exchanges. For column (2), we perform the regression

$$buy_{x,t} = \beta_0 + \beta_1(a_{-x,t} - a_{x,t}) + \beta_2(a_{x,t} - v_t) + \varepsilon_{x,t},$$

and similarly, for column (3), we perform the regression

$$sell_{x,t} = \beta_0 + \beta_1(b_{x,t} - b_{-x,t}) + \beta_2(v_t - b_{x,t}) + \varepsilon_{x,t}.$$

In both cases, the sample consists of all exchanges $x \in \{\text{ASX}, \text{Chi-X}\}$ and all seconds t between 10:30 and 16:00 in the 80 trading days of the sample. Moreover, because we do not observe v_t, we proxy with the average cum-fee mid price $(b_{\text{ASX},t} + b_{\text{Chi-X},t} + a_{\text{ASX},t} + a_{\text{Chi-X},t})/4$ in both cases. In the table, we refer to the difference in the own and other cum-fee asks (or the own and other cum-fee bids) as the "price difference," and we refer to the difference between the own cum-fee ask and v_t (or the own cum-fee bid and v_t) as the "half spread."

In addition, column (1) of the table reports the results of combining both regressions and estimating them together. Formally, we index the sides of the trade by $i \in \{\text{buy}, \text{sell}\}$. We then define $iso_{\text{buy},x,t} = buy_{x,t}$ and $iso_{\text{sell},x,t} = sell_{x,t}$. Likewise, we define $price\_difference_{\text{buy},x,t} = a_{-x,t} - a_{x,t}$ and $price\_difference_{\text{sell},x,t} = b_{x,t} - b_{-x,t}$. Finally, we define $half\_spread_{\text{buy},x,t} = a_{x,t} - v_t$ and $half\_spread_{\text{sell},x,t} = v_t - b_{x,t}$. Given these definitions, column (1) reports the results of the regression

$$iso_{i,x,t} = \beta_0 + \beta_1 price\_difference_{i,x,t} + \beta_2 half\_spread_{i,x,t} + \varepsilon_{i,x,t},$$

where the sample consists of both sides $i \in \{\text{buy}, \text{sell}\}$, all exchanges $x \in \{\text{ASX}, \text{Chi-X}\}$, and all seconds t between 10:30 and 16:00 in the 80 trading days of the sample. Moreover, we proxy for v_t with the average cum-fee mid price, as before.

Focusing on the estimates in column (1), we find that a one cent increase (decrease) in the ask (bid) on an exchange is associated with a decrease in the number of isolated buys (sells) on that exchange of approximately 1.5 per hour, as well as an increase in the number of isolated buys (sells) at the other exchange of approximately 2.3 per hour. Qualitatively similar results are found in column (2), where the focus is only on isolated buys, and in column (3), where the focus is only on isolated sells.

Table 14: Isolated Trades as Function of Quotes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. For columns (2) and (3), an observation is an exchange-second pair, where the exchange is either ASX or Chi-X, and the second is between 10:30 and 16:00 in one of the 80 trading days in the sample. The dependent variables are, respectively, indicators for the occurrence of an isolated buy or an isolated sell on that exchange in that second. For column (1), the estimation combines buys and sells. The independent variable "price difference" is defined as the difference between the own and other cum-fee bids or asks (see appendix E.C). The independent variable "half spread" is defined as the difference between the own cum-fee bid or ask and the mid price (see appendix E.C). Coefficients are estimated by ordinary least squares. Standard errors are clustered by 120 second blocks on each trading day.

	(1) BUY or SELL	(2) BUY	(3) SELL
	(1)	(2)	(3)
price difference	0.000638***	0.000548***	0.000737***
	(0.0000566)	(0.0000657)	(0.0000772)
half spread	-0.000405***	-0.000379***	-0.000431***
	(0.000103)	(0.000111)	(0.000133)
Constant	0.00291***	0.00293***	0.00290***
	(0.000149)	(0.000163)	(0.000193)
Observations	6336000	3168000	3168000

E.D Trade Clustering and Price Changes

Our estimation strategy relies upon using isolated trades as a proxy for the liquidity-motivated investor trades of the model. Conversely, clustered trades are interpreted as the information-motivated sniper trades of the model. If this approach is valid, then we should expect clustered trades to be better predictors of subsequent price movements than their isolated counterparts. In this appendix, we present evidence to show that this relationship is indeed borne out in the data.

To that end, we define $R(\Delta)_t = (m_{t+\Delta} - m_t)/m_t$ to be the return over a time interval of length Δ, where m_t is the NBBO mid price. We define indicators $clusterBuy_t$ and $clusterSell_t$ for, respectively, whether a clustered buy or a clustered sell occurs on either exchange in second t. We also define $B_t = buy_{\text{ASX},t} + buy_{\text{Chi-X},t} + clusterBuy_t$ and $S_t = sell_{\text{ASX},t} + sell_{\text{Chi-X},t} + clusterSell_t$ to be indicators for, respectively, whether a buy or a sell (in each case, either clustered or isolated) occurs on either exchange in second t.

Then, for each value of Δ, we run the regression

$$R(\Delta)_t = \alpha_\Delta + \beta_{1,\Delta} B_t + \beta_{2,\Delta}\, clusterBuy_t + \gamma_{1,\Delta} S_t + \gamma_{2,\Delta}\, clusterSell_t + \varepsilon_t.$$

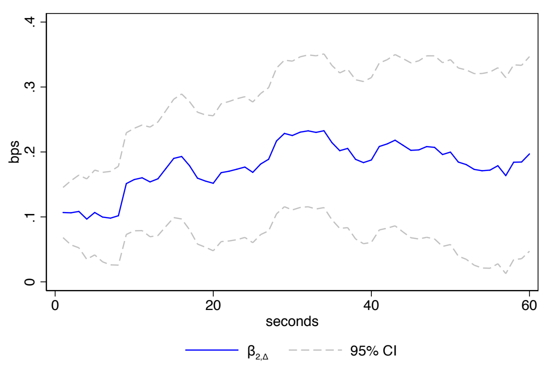
In each case, the sample consists of all seconds t in the 80 trading days of the sample that are between (*i*) 10:30 and (*ii*) Δ seconds before 16:00.

In Figure 5, we plot point estimates and 95% confidence intervals for $\beta_{2,\Delta}$ and $\gamma_{2,\Delta}$ for values of Δ up to one minute. As expected, we find that clustered trades are more predictive of price movements. The occurrence of a clustered buy (sell) in second t is associated with a 30-second return of 0.23 bps (−0.21 bps) more than an isolated buy (sell). Qualitatively, the result is robust to the choice of time horizon.

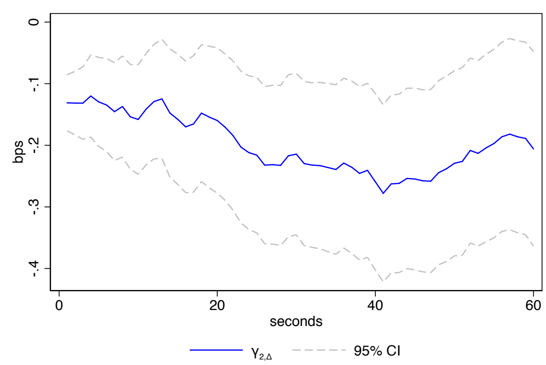
These findings are also consistent with those of Menkveld (2018), who similarly finds that trades arriving in clusters exert a disproportionately high amount of adverse selection.

Figure 5: Predicting Price Changes Using Trade Clustering

The graphs plot point estimates for $\beta_{2,\Delta}$ and $\gamma_{2,\Delta}$ from the regression defined in the text for Δ ranging from 1 to 60 seconds. An observation is a second between (*i*) 10:30 and (*ii*) Δ seconds before 16:00 in one of the 80 trading days in the sample. 95% confidence intervals as based on standard errors clustered by 120 second blocks on each trading day.



(a) Aggressive BUY



(b) Aggressive SELL

F Extensions of the Model

Several of the assumptions embedded in the baseline model are only for ease of exposition. This appendix discusses seven extensions of the model, under each of which all our main results extend. In appendix F.A, we allow for the possibility that each HFT faces a constraint on its net inventory. In appendix F.B, we add costs of operation for exchanges and liquidity providers. In appendix F.C, we consider an extension of the model in which the process governing the evolution of the value of the security is enriched to include heterogeneous jump sizes and a Brownian component. In appendix F.D, we relax the indivisibility constraint on investor demand, allowing them to split orders across exchanges. In appendix F.E, we allow for some agents who acquire and trade on short-lived private information. In appendix F.F, we allow for the possibility that liquidity providers win an exogenous fraction of races to respond to jumps in the value of the security. Finally, in appendix F.G, we allow the horizon T to be stochastic. None of these extensions affects the expressions for the equilibrium spread. Consequently, our empirical findings remain valid under each.

Lastly, in appendix F.H, we explain how an imperfect ability to monitor prices in real time might produce market frictions of the nature assumed in the baseline model.

F.A Inventory Constraints

In the baseline model, HFTs face no limits on the amount of inventory that they are able to take on. This is admittedly somewhat unrealistic. In contrast, due to capital controls and a desire to limit their risk, HFTs typically limit the extent to which they allow themselves to build up large inventories.

While a rich literature in market microstructure theory studies market maker inventory management, most of those papers feature a monopolistic market maker (Stoll, 1978; Amihud and Mendelson, 1980; Ho and Stoll, 1981). In contrast, our model features an infinite number of HFTs, any one of which can provide liquidity. As we argue here, inventory management is a less important determinant of the equilibrium spread when market making is competitive in this way. Under the additional assumption that any HFT can take on a small amount of inventory without cost, the equilibrium spread remains unchanged.

Suppose that each HFT faces inventory constraints that prohibit its net position from exceeding $K \geq 1$ shares long or short at any point in time.[54] Given this modification of the model, it is relatively straightforward to adapt the proofs of propositions 1 and 2 to show that the equilibrium spread remains unchanged. The strategies of HFTs adjust so that they become inactive if their net inventory ever reaches $\pm(K-1)$ shares long or short. And if the liquidity provider ever becomes inactive, then one of the snipers abandons its original strategy to assume the liquidity provider's role.

F.B Cost of Operation

In the baseline model, neither exchanges nor traders must pay a cost of operation. In this extension, we add per-time costs of operation for exchanges and liquidity providers. Although one might have suspected that these operating costs would be an additional source of a spread, they affect neither the equilibrium spread nor therefore our empirical findings.

Formally, suppose that an HFT who has active limit orders on x exchanges must pay a monitoring cost of xc_1. Suppose also that an operation cost of c_2 must be paid by any active exchange. Also, amend A3 to read:

A3′. $\lambda_i \left(1 - \frac{1}{\theta}\frac{\Sigma}{2}\right)\frac{\Sigma}{2} \geq \lambda_j X \left(\sigma - \frac{\Sigma}{2}\right) + X(c_1 + c_2)$

Given the modified assumption, it is relatively straightforward to extend the proofs of propositions 1 and 2 to show that the equilibrium spread remains unchanged. The strategies of the liquidity providers adjust to accommodate the inclusion of c_1 in their zero-profit conditions. Working backwards from this, it can be shown that the profits of each exchange,

[54] This specification can be thought of as a special case of a model in which HFTs face a cost of inventory that is a function of their net position (which in itself can be thought of as reduced form for a model in which HFTs are risk averse). Specifically, this corresponds to the special case in which the inventory cost is zero on $[-K, K]$ and prohibitively high outside that interval.

as a function of the cum-fee spreads, are $c_1 + c_2$ less than before. These additive constants fall out of the optimization problem, and the same spreads arise in equilibrium.

There are other costs (e.g., order processing costs and inventory costs) for which the analysis is less clean. Nevertheless, the findings discussed above provide us with some confidence that our empirical findings would not necessarily be diluted or reversed by enriching the model to include various costs, even those commonly interpreted as components of the spread.

F.C Enriched Information Arrival Processes

In the baseline model, v_t is affected by jumps of only one size, σ, which arrive at the rate λ_j. While this is a quite simple model of price movement, the main conclusions remain unchanged even if the process is enriched in various ways.

In particular, the findings remain intact if the baseline jump process is augmented with a Brownian motion, possibly with time-varying volatility. Equilibrium strategies change only slightly under this extension: liquidity providers update their stale quotes not only after jumps, but also in a continuous fashion that tracks the Brownian motion. Snipers, on the other hand, attempt to trade only upon jumps. Given this, it is relatively straightforward to establish that the equilibrium spread continues to be as described by propositions 1 and 2.

Furthermore, the findings would also remain unchanged if the size of each discrete jump were drawn independently from a distribution, F that satisfies the following modifications of A2 and A3:

A2″. $\theta < \min \operatorname{supp}(F)$.

A3″. $\lambda_i \left(1 - \frac{1}{\theta}\frac{\Sigma}{2}\right)\frac{\Sigma}{2} \geq \lambda_j X \left(\int \sigma dF(\sigma) - \frac{\Sigma}{2}\right)$, where Σ, as before, is defined in terms of the underlying parameters as follows:

$$\Sigma \equiv \begin{cases} \theta\left(1 + \dfrac{\lambda_j}{\lambda_i}\right) & \text{if } X = 1 \\ \theta + \dfrac{4\alpha}{X^2} - \sqrt{\theta^2 + \dfrac{16\alpha^2}{X^4} - \dfrac{8\alpha\theta\lambda_j}{X\lambda_i}} & \text{if } X \geq 2 \end{cases}$$

In words, A2 is modified to apply to the minimum jump size, and A3 is modified to apply to the expected jump size.

F.D Order Splitting

In the baseline model, investor demand is indivisible: each investor is limited to buying or selling a single share at a single exchange. In this extension, we allow investors to split orders across exchanges in such a way that they buy or sell one share in total. The baseline equilibrium survives unchanged even if this form of order splitting is allowed.[55]

Suppose that an investor arriving at time t has the following available actions: (i) submit

[55]However, we have not ruled out the possibility that this modification might introduce new equilibria.

a profile of market orders $\boldsymbol{y} = (y_1, \ldots, y_X)$, where $y_x \geq 0$ for each exchange x and $\sum_{x=1}^{X} y_x = 1$; (*ii*) submit a profile of market orders $\boldsymbol{y} = (y_1, \ldots, y_X)$, where $y_x \leq 0$ for each exchange x and $\sum_{x=1}^{X} y_x = -1$; or (*iii*) submit a profile of market orders $\boldsymbol{y} = (y_1, \ldots, y_X)$, where $y_x = 0$ for each exchange x.

Supposing that liquidity providers continue to quote one share at the bid and one share at the ask, then an investor who arrives at time t and submits the orders $\boldsymbol{y}$ obtains utility

$$u_t(\boldsymbol{y}|\tilde{\theta}) = \begin{cases} v_t + \tilde{\theta} - \sum_{x=1}^{X} y_x a_{x,t} & \text{if } \sum_{x=1}^{X} y_x = 1 \\ \sum_{x=1}^{X} |y_x| b_{x,t} - v_t - \tilde{\theta} & \text{if } \sum_{x=1}^{X} y_x = -1 \\ 0 & \text{if } \sum_{x=1}^{X} y_x = 0 \end{cases}$$

As in the baseline model, investors do not necessarily act to maximize their utility. Instead, an investor who arrives at time t chooses orders $\boldsymbol{y}$ to maximize

$$\hat{u}_t(\boldsymbol{y}|\tilde{l}, \tilde{\theta}) = u_t(\boldsymbol{y}|\tilde{\theta}) - 2\alpha \cdot \sum_{x=1}^{X} |y_x| d(\tilde{l}, l_x)^2.$$

Note that conditional on choosing $\boldsymbol{y}$ so that $\sum_{x=1}^{X} y_x = 1$, it is optimal under $\hat{u}$ to trade the entire quantity at the exchange x that maximizes $-a_{x,t} - 2\alpha d(\tilde{l}, l_x)^2$. Likewise, conditional on choosing $\boldsymbol{y}$ so that $\sum_{x=1}^{X} y_x = -1$, it is optimal under $\hat{u}$ to trade the entire quantity at the exchange x that maximizes $b_{x,t} - 2\alpha d(\tilde{l}, l_x)^2$. In other words, the investor does not avail himself of the opportunity to split orders, as a result of the linear way in which $|y_x|$ enters $\hat{u}$, and his behavior remains as in the baseline model.

To show that the baseline equilibrium remains intact under this modification, it remains only to verify that no liquidity provider can profitably deviate by quoting less than one share of depth. Reducing depth in this way would reduce the volume of trade with snipers, but it would also induce investors who would have traded at that exchange to reduce their demand by a proportionate amount. This leaves the liquidity provider's zero-profit condition unchanged, and so the deviation would not be profitable.

F.E Private Information

In the baseline model, information is purely public. Nevertheless, the main conclusions remain unchanged even if short-lived private information is incorporated in a particular way.

In this extension, we augment the model by adding a trader ("the analyst") who may acquire private information. Jumps in the value of the asset continue to be of size σ and arrive at the rate λ_j. A fraction η of the jumps are, as in the baseline model, revealed publicly when they occur. The remaining $1 - \eta$ fraction of the jumps are revealed publicly only after an infinitesimal delay. But the analyst observes *every* jump when it occurs, and obtains short-lived private information in the latter cases. The analyst may trade at as many exchanges as he wishes, but is restricted to immediate-or-cancel orders.

In the equilibrium of this extension, the analyst submits immediate-or-cancel orders to each exchange to buy (sell) one share immediately after observing an upward (downward) jump. Equilibrium strategies of the liquidity providers change only slightly in this extension: after a trade against one of their orders, they wait for an infinitesimal length of time before replenishing the limit order book, updating the price if information arrives in the interim. Equilibrium strategies of investors and snipers remain unchanged. Given this, it is relatively straightforward to show that the expressions for the equilibrium spreads prevailing under the limit order book remain as before.

There would, however, be changes to the expressions for the spreads prevailing under frequent batch auctions and non-cancellation delays (*cf.* appendix C.A). Intuitively, these alternative trading mechanisms remove the adverse selection due to stale-quote sniping on the basis of public information, but not the adverse selection due to the analyst's private information. Thus, while these mechanisms would continue to improve upon the limit order book in the presence of private information of this nature, the reduction in transaction costs would be less dramatic.

F.F Incorporating Quote Fade

In the baseline model, jumps in the value of the security set off races between the liquidity providers, attempting to cancel their mispriced quotes, and snipers, attempting to exploit them. For the reason that each liquidity provider races against an infinite number of snipers, the liquidity providers lose their races with probability one.

In contrast, liquidity providers are sometimes successful in cancelling mispriced quotes in practice, a phenomenon that is sometimes referred to as quote fade. This might be for a number of unmodelled reasons, including: (*i*) there are only a finite number of snipers in practice, (*ii*) snipers might be less strongly incentivized to monitor the market and/or invest in speed technology than liquidity providers, or (*iii*) snipers might require stronger signals to react than liquidity providers do.[56]

Without formally modeling any of the aforementioned reasons, one way to modify our model so as to feature quote fade would be to assume that an exogenous fraction $\phi \in [0,1)$ of the races are won by liquidity providers. All our findings would remain unchanged. Formally, we would reinterpret the parameter λ_j as $(1-\phi)\lambda'_j$, where λ'_j now denotes the underlying arrival rate of jumps in the value of the security. And λ_j, instead of representing the arrival rate of jumps, now represents that rate scaled by the fraction of races that the liquidity provider loses. Given this reinterpretation, it is straightforward to show that the expressions for the equilibrium spreads remain as before.

Moreover, with this modification our model seems to us a reasonably close approximation of alternate versions of the model that would capture—in a more formal way—the aforementioned explanations for quote fade. We therefore speculate that such versions of the model would produce results that would be qualitatively similar to our current conclusions (although perhaps not exactly the same).

[56] In Baldauf and Mollner (forthcoming), we formally demonstrate how this can arise (albeit with a model in which adverse selection originates from private as opposed to public information).

F.G Stochastic Terminal Time

The baseline model assumes a deterministic, finite horizon. However, this is only for simplicity of exposition. Our analysis immediately generalizes to cases in which the terminal time T is stochastic. The reason is that, although the model is set in continuous time, dynamics play no role: equilibrium trading behavior is stationary, just as in Budish et al. (2015).

F.H Market Friction Microfoundation

A key aspect of the model is that it allows for market frictions that distort the exchange choices of investors. In the main text, we are deliberately agnostic as to the precise source of these frictions. Nevertheless, we do suggest a number of potential sources, one of which is that it is inherently difficult to monitor prices in real time with perfect accuracy. In this appendix, we develop this explanation in greater detail.

For the purposes of this appendix, we focus on the case in which $X = 2$. As in the main text, denote the private transaction motive of an investor by $\tilde{\theta}$. Ignore, for the purposes of this appendix, the other component of the investor's type, $\tilde{l}$, which had designated the investor's location on the unit circle. For expositional ease, we focus the following discussion on investors with an inclination to buy (i.e., for whom $\tilde{\theta} \geq 0$) and, correspondingly, the ask-side of the book. Denote the asks at the two exchanges at time t by $a_{1,t}$ and $a_{2,t}$. Suppose, however, that investors are limited in their capacity to monitor these prices. This could be the case if their information comes from some lagged feed (such as that of the Securities Information Processor, often abbreviated as "the SIP"), or if so-called "fleeting orders" prevent investors from obtaining a clear view of the market.

As a stylized model of these limitations, suppose that rather than observing the two asks, an investor who arrives at time t observes only the noisy signals $\hat{a}_{1,t} = a_{1,t} + \varepsilon_t$ and $\hat{a}_{2,t} = a_{2,t} - \varepsilon_t$, where $\varepsilon_t \sim U[-\alpha, \alpha]$. Suppose that the investor routes an immediate-or-cancel order with limit price $v_t + \tilde{\theta}$ to the exchange with best (i.e., lowest) signal.[57]

This behavior induces the same trading probabilities as would be obtained if investors of every type $(\tilde{l}, \tilde{\theta})$ were to optimize $\hat{u}_t(x, y|\tilde{l}, \tilde{\theta})$ as described in the main text, and we were then to integrate over $\tilde{l}$. In particular, conditioning on the true quotes $(a_{1,t}, a_{2,t})$, the investor routes to exchange 1 with probability $\left[\frac{1}{2} + \frac{a_{2,t} - a_{1,t}}{\alpha}\right]_0^1$. Then, given the limit price that the investor sets, this means that the investor trades at exchange 1 with probability $\left[\frac{1}{2} + \frac{a_{2,t} - a_{1,t}}{\alpha}\right]_0^1 \mathbb{1}\left\{v_t + \tilde{\theta} \geq a_{1,t}\right\}$. And we obtain a symmetric expression for exchange 2.

[57] Such behavior is optimal for the investor if he is constrained to send orders only at time t and if he possesses prior beliefs about ask prices that are symmetric about $a_{1,t}$ and $a_{2,t}$.

G Notation and Variables

Table 15: List of Mathematical Notation

Name	Description
Parameters	
α	strength of frictions distorting exchange choice of investors
θ	maximum investor private transaction motive
λ_i	Poisson arrival rate of investors
λ_j	Poisson arrival rate of jumps
σ	jump size
X	number of exchanges
Other Notation	
v	fundamental value of security
s	cum-fee spread*
a	cum-fee ask*
b	cum-fee bid*
τ_{make}	make fee
τ_{take}	take fee
l	location on unit circle

We also use $\hat{s}$, $\hat{a}$, and $\hat{b}$ for the quoted spread, ask, and bid, respectively.

Table 16: Variables Used in Main Estimation

Name	Description
$s_{x,t}$	cum-fee spread of STW on exchange x in second t (cents)
$a_{x,t}$	cum-fee ask of STW on exchange x in second t (cents)
$b_{x,t}$	cum-fee bid of STW on exchange x in second t (cents)
v_t	fundamental value of security*
$buy_{x,t}$	indicator for an isolated buy of STW on exchange x in second t
$sell_{x,t}$	indicator for an isolated sell of STW on exchange x in second t
$clustered_{x,t}$	indicator for a clustered trade of STW on exchange x in second t

Proxied by $(a_{\text{ASX,t}} + b_{\text{ASX,t}} + a_{\text{Chi-X,t}} + b_{\text{Chi-X,t}})/4$.

References for Internet Appendix

Aitken, Michael, Haoming Chen, and Sean Foley, "The Impact of Fragmentation, Exchange Fees and Liquidity Provision on Market Quality," *Journal of Empirical Finance*, 2017, *41*, 140–160.

Amihud, Yakov and Haim Mendelson, "Dealership Market: Market-Making with Inventory," *Journal of Financial Economics*, 1980, *8* (1), 31–53.

_ , **Beni Lauterbach, and Haim Mendelson**, "The Value of Trading Consolidation: Evidence From the Exercise of Warrants," *Journal of Financial and Quantitative Analysis*, 2003, *38* (4), 829–846.

Arnold, Tom, Philip Hersch, J. Harold Mulherin, and Jeffry Netter, "Merging Markets," *The Journal of Finance*, 1999, *54* (3), 1083–1107.

ASX Group, "ASX Centre Point Block," 2012. http://www.asx.com.au/documents/products/asx_centre_point_factsheet.pdf.

_ , "ITCH - Glimpse Message Specification," ASX Market Information 2012.

_ , "ASX Operating Rules Procedures," 2016. http://www.asx.com.au/documents/rules/asx_or_procedures.pdf.

_ , "Managed Funds and ETP Product List," 2017. http://www.asx.com.au/products/etf/managed-funds-etp-product-list.htm.

Australian Securities and Investments Commission, "ASIC Market Integrity Rules (Competition in Exchange Markets)," 2011. http://asic.gov.au/regulatory-resources/markets/market-integrity-rules.

_ , "Report 452: Review of High-Frequency Trading and Dark Liquidity," 2015. http://download.asic.gov.au/media/3444836/rep452-published-26-october-2015.pdf.

_ , "List of Crossing Systems Registered with ASIC," 2016. http://www.asic.gov.au/crossing-systems.

Baldauf, Markus and Joshua Mollner, "High-Frequency Trading and Market Performance," *The Journal of Finance*, forthcoming. http://ssrn.com/abstract=2674767.

Battalio, Robert H., "Third Market Broker-Dealers: Cost Competitors or Cream Skimmers?," *The Journal of Finance*, 1997, *52* (1), 341–352.

Bennett, Paul and Li Wei, "Market Structure, Fragmentation, and Market Quality," *Journal of Financial Markets*, 2006, *9* (1), 49–78.

Bernales, Alejandro, Italo Riarte, Satchit Sagade, Marcela Valenzuela, and Christian Westheide, "A Tale of One Exchange and Two Order Books: Effects of Fragmentation in the Absence of Competition," *Working Paper*, 2017.

Bessembinder, Hendrik and Herbert M. Kaufman, "A Cross-Exchange Comparison of Execution Costs and Information Flow for NYSE-Listed Stocks," *Journal of Financial Economics*, 1997, *46* (3), 293–319.

Biais, Bruno, Christophe Bisière, and Chester Spatt, "Imperfect Competition in Financial Markets: An Empirical Study of Island and Nasdaq," *Management Science*, 2010, *56* (12), 2237–2250.

Boehmer, Beatrice and Ekkehart Boehmer, "Trading Your Neighbor's ETFs: Competition or Fragmentation?," *Journal of Banking & Finance*, 2003, *27* (9), 1667–1703.

Boneva, Lena, Oliver Linton, and Michael Vogt, "The Effect of Fragmentation in Trading on Market Quality in the UK Equity Market," *Journal of Applied Econometrics*, 2016, *31* (1), 192–213.

Branch, Ben and Walter Freed, "Bid-Asked Spreads on The AMEX and The Big Board," *The Journal of Finance*, 1977, *32* (1), 159–163.

Budish, Eric, Peter Cramton, and John Shim, "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," *The Quarterly Journal of Economics*, 2015, *130* (4), 1547–1621.

Cameron, A. Colin and Pravin K. Trivedi, *Microeconometrics: Methods and Applications*, Cambridge University Press, 2005.

Chi-X Australia, "Market Data Feed Specification," Chi-X Global Inc. 2012.

_ , "Operating Rules: Procedures," 2013. `https://www.chi-x.com.au/wp-content/uploads/2017/02/CXA-Operating-Rules-Procedures-v1.4.pdf`.

Chlistalla, Michael and Marco Lutat, "Competition in Securities Markets: The Impact on Liquidity," *Financial Markets and Portfolio Management*, 2011, *25* (2), 149–172.

Cohen, Kalman J. and Robert M. Conroy, "An Empirical Study of the Effect of Rule 19c-3," *Journal of Law and Economics*, 1990, *33* (1), 277–305.

Comerton-Forde, Carole and Tālis J. Putniņš, "Dark Trading and Price Discovery," *Journal of Financial Economics*, 2015, *118* (1), 70–92.

Degryse, Hans, Frank de Jong, and Vincent van Kervel, "The Impact of Dark Trading and Visible Fragmentation on Market Quality," *Review of Finance*, 2015, *19* (4), 1587–1622.

Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph B. Paperman, "Liquidity, Information, and Infrequently Traded Stocks," *The Journal of Finance*, 1996, *51* (4), 1405–1436.

_ **, Robert F. Engle, Maureen O'Hara, and Liuren Wu**, "Time-Varying Arrival Rates of Informed and Uninformed Trades," *Journal of Financial Econometrics*, 2008, *6* (2), 171–207.

Fink, Jason, Kristin E. Fink, and James P. Weston, "Competition on the Nasdaq and the Growth of Electronic Communication Networks," *Journal of Banking & Finance*, 2006, *30* (9), 2537–2559.

Foley, Sean and Tālis J. Putniņš, "Should We be Afraid of the Dark? Dark Trading and Market Quality," *Journal of Financial Economics*, 2016, *122* (3), 456–481.

Fontnouvelle, Patrick De, Raymond P.H. Fishe, and Jeffrey H. Harris, "The Behavior of Bid-Ask Spreads and Volume in Options Markets During the Competition for Listings in 1999," *The Journal of Finance*, 2003, *58* (6), 2437–2464.

Foucault, Thierry and Albert J. Menkveld, "Competition for Order Flow and Smart Order Routing Systems," *The Journal of Finance*, 2008, *63* (1), 119–158.

Gajewski, Jean-François and Carole Gresse, "Centralised Order Books Versus Hybrid Order Books: A Paired Comparison of Trading Costs on NSC (Euronext Paris) and SETS (London Stock Exchange)," *Journal of Banking & Finance*, 2007, *31* (9), 2906–2924.

Gill, Philip E., Elizabeth Wong, Walter Murray, and Michael A. Saunders, "User's Guide for SNOPT Version 7.5: Software for Large-Scale Nonlinear Programming," 2015. `http://www.cam.ucsd.edu/~peg/papers/sndoc7.pdf`.

Glosten, Lawrence R. and Lawrence E. Harris, "Estimating the Components of the Bid/Ask Spread," *Journal of Financial Economics*, 1988, *21* (1), 123–142.

Hall, Peter, Joel L. Horowitz, and Bing-Yi Jing, "On Blocking Rules For the Bootstrap with Dependent Data," *Biometrika*, 1995, *82* (3), 561–574.

Hamilton, James L., "Marketplace Fragmentation, Competition, and the Efficiency of the Stock Exchange," *The Journal of Finance*, 1979, *34* (1), 171–187.

Haslag, Peter H. and Matthew Ringgenberg, "The Demise of the NYSE and NASDAQ: Market Quality in the Age of Market Fragmentation," *Working Paper*, 2017. `http://ssrn.com/abstract=2591715`.

He, Peng William, Elvis Jarnecic, and Yubo Liu, "The Determinants of Alternative Trading Venue Market Share: Global Evidence from the Introduction of Chi-X," *Journal of Financial Markets*, 2015, *22*, 27–49.

Hendershott, Terrence and Charles M. Jones, "Island Goes Dark: Transparency, Fragmentation, and Regulation," *The Review of Financial Studies*, 2005, *18* (3), 743–793.

Ho, Thomas and Hans R. Stoll, "Optimal Dealer Pricing Under Transactions and Return Uncertainty," *Journal of Financial Economics*, 1981, *9* (1), 47–73.

Mayhew, Stewart, "Competition, Market Structure, and Bid-Ask Spreads in Stock Option Markets," *The Journal of Finance*, 2002, *57* (2), 931–958.

Menkveld, Albert J., "High Frequency Trading and the *New Market* Makers," *Journal of Financial Markets*, 2013, *16* (4), 712–740.

_ , "High-Frequency Trading as Viewed Through an Electron Microscope," *Financial Analysts Journal*, 2018, *74* (2), 24–31.

Neal, Robert, "Potential Competition and Actual Competition in Equity Options," *The Journal of Finance*, 1987, *42* (3), 511–531.

Nguyen, Vanthuan, Bonnie F. Van Ness, and Robert A. Van Ness, "Short-and Long-Term Effects of Multimarket Trading," *Financial Review*, 2007, *42* (3), 349–372.

Nielsson, Ulf, "Stock Exchange Merger and Liquidity: The Case of Euronext," *Journal of Financial Markets*, 2009, *12* (2), 229–267.

O'Hara, Maureen and Mao Ye, "Is Market Fragmentation Harming Market Quality?," *Journal of Financial Economics*, 2011, *100* (3), 459–474.

Sandås, Patrik, "Adverse Selection and Competitive Market Making: Empirical Evidence from a Limit Order Market," *The Review of Financial Studies*, 2001, *14* (3), 705–734.

Stoll, Hans R., "The Supply of Dealer Services in Securities Markets," *The Journal of Finance*, 1978, *33* (4), 1133–1151.

Weston, James P., "Electronic Communication Networks and Liquidity on the Nasdaq," *Journal of Financial Services Research*, 2002, *22* (1/2), 125–139.